

Air Force Institute of Technology

AFIT Scholar

Theses and Dissertations

Student Graduate Works

3-2007

Sampled-Data Kalman Filtering and Multiple Model Adaptive Estimation for Infinite-Dimensional Continuous-Time Systems

Scott A. Sallberg

Follow this and additional works at: <https://scholar.afit.edu/etd>



Part of the [Controls and Control Theory Commons](#)

Recommended Citation

Sallberg, Scott A., "Sampled-Data Kalman Filtering and Multiple Model Adaptive Estimation for Infinite-Dimensional Continuous-Time Systems" (2007). *Theses and Dissertations*. 2888.

<https://scholar.afit.edu/etd/2888>

This Dissertation is brought to you for free and open access by the Student Graduate Works at AFIT Scholar. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of AFIT Scholar. For more information, please contact richard.mansfield@afit.edu.

AFIT/DS/ENG/07-08



SAMPLED-DATA KALMAN FILTERING AND
MULTIPLE MODEL ADAPTIVE ESTIMATION
FOR INFINITE-DIMENSIONAL
CONTINUOUS-TIME SYSTEMS

DISSERTATION
Scott A. Sallberg, Major, USAF

AFIT/DS/ENG/07-08

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

Approved for public release; distribution unlimited.

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

AFIT/DS/ENG/07-08

SAMPLED-DATA KALMAN FILTERING AND
MULTIPLE MODEL ADAPTIVE ESTIMATION FOR
INFINITE-DIMENSIONAL CONTINUOUS-TIME SYSTEMS

DISSERTATION

Presented to the Faculty of the
Graduate School of Engineering and Management
Air Force Institute of Technology
Air University
Air Education and Training Command
In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Scott A. Sallberg, B.S.E.E., M.S.E.E.
Major, USAF

March 2007

Approved for public release; distribution unlimited.

SAMPLED-DATA KALMAN FILTERING AND
MULTIPLE MODEL ADAPTIVE ESTIMATION FOR
INFINITE-DIMENSIONAL CONTINUOUS-TIME SYSTEMS

Scott A. Sallberg, B.S.E.E., M.S.E.E.

Major, USAF

Approved:

<u>Peter S. Maybeck</u>	<u>28 Feb 2007</u>
Peter S. Maybeck, Ph.D.	Date
Research Committee Chairman	

<u>Mark E. Oxley</u>	<u>28 Feb 2007</u>
Mark E. Oxley, Ph.D.	Date
Research Committee Member	

<u>M. Pachter</u>	<u>28 Feb. 2007</u>
Meir Pachter, Ph.D.	Date
Research Committee Member	

<u>William P. Baker</u>	<u>28 Feb 07</u>
William P. Baker, Ph.D.	Date
Dean's Representative	

Accepted:

<u>M U Thomas</u>
M. U. Thomas, Ph.D.
Dean, Graduate School of Engineering and Management

Acknowledgments

During the course of this six-and-a-half-year academic odyssey, my determination to complete the dissertation document has waxed and waned as unforeseen difficulties repeatedly arose these last few years. My hope for a quick and successful completion of this research document has been dashed more than once. In his short book on heuristics, Dr. George Polya wrote that [162]

It would be a mistake to think that solving problems is a purely “intellectual affair”; determination and emotions play an important role. Lukewarm determination and sleepy consent to do a little something may be enough for a routine problem in the classroom. But, to solve a serious scientific problem, will power is needed that can outlast years of toil and bitter disappointments.

While I was responsible for the content of this research, it would be sheer folly to think that I did this alone; numerous people have guided, assisted, encouraged, supported, inspired, and humored me while I researched, studied, derived, and composed this doctoral dissertation on Kalman filtering. The first round of thanks goes to my research committee: Dr. Peter Maybeck, Dr. Mark Oxley, Dr. Meir Pachter, and Dr. William Baker. The grandiose ideas that I presented at my prospectus defense more than four years ago have been refined and transformed into a completed dissertation because of their sage advice and expert guidance.

Even though it has been more than five years since I attended Prof Maybeck’s three-part course on stochastic estimation and control, his knowledge of and passion for the subject matter is still alive in my mind today. Thank you Prof Maybeck for chairing my research committee, patiently guiding me all these years, and for reviewing a never-ending stream of draft chapters. In a similar vein, Prof Oxley enlivened the mathematics of functional analysis — one of the main mathematical tools used in this research. Thank you Prof Oxley for the endless hours of math talk, for bringing out the mathematician in this engineer, and for refining my imprecise mathematical prose.

While the professors occupied center stage during my formal education, the daily interaction with fellow students in and out of the classroom was invaluable. Thank you Jesse Zydallis, Eddie Ochoa, Vern Starman, Ken Pascoe, Jeff Hebert, Chuck Ormsby, Matt Campbell, Virgil Zetterlind, Nick Wager, Pat Kee, Rick Day, Lance Champaign, Stacie Taylor, Nate Lockwood, Jay Cucko, Ron Coutu, Jeremy Raley, Ken Fisher, Steve Cox, Dan Gisselquist, Steve Thorsen, John Seo, Doug Decker, Mike Winthrop, Dan Millman, Phil Corbell, Dave Parker, Pete Crabtree, Jon Luminati, and Eric Nelson for your conversation, homework assistance/collaboration, and/or sound counsel. You all have immeasurably enriched my life and studies.

A special thanks goes out to the folks who created MATLAB, L^AT_EX, and WinEdt — without the benefit of these incredible software tools that I used to simulate and document this research, ten years would not have been enough time to complete this journey.

Ever since I was a wee small lad, my favorite retort to nearly any statement has always been *why?* Thanks Mom and Dad for “encouraging” me to ask questions and to seek out the *why?* for myself — I love you guys.

My family deserves much of the credit for this work. It was only through the love and support of my wife and children that I was able to complete this dissertation document; I could not have done it without them. Thank you Kemmi, I love you.

During the course of this arduous and incredibly rewarding journey, I have learned that with faith, perseverance, hope, and God, that all things are possible if you are surrounded by good people.

Scott A. Sallberg

Table of Contents

	Page
Acknowledgments	iii
List of Figures	xi
List of Tables	xvi
List of Symbols	xvii
List of Abbreviations	xxvi
Abstract	xxviii
 I. Introduction	 1-1
1.1 Overview	1-3
1.2 Historical Overview	1-6
1.3 Example: Navigation	1-8
1.3.1 Kalman Filter with Known Noise Strengths	1-8
1.3.2 Multiple Model Adaptive Estimation	1-15
1.4 Advanced Topics Overview	1-22
1.5 Notes to the Reader	1-23
1.6 Summary	1-26
1.7 The Rest of the Dissertation	1-26
 II. Multiple Model Adaptive Estimation	 2-1
2.1 Introduction	2-1
2.2 The Beginnings of Multiple Model Adaptive Estimation and Control	2-2
2.3 Multiple Model Adaptive Estimation Fundamentals	2-7

	Page
2.3.1 Mathematical System Model	2-8
2.3.2 Kalman Filtering	2-14
2.3.3 Filter Bank	2-17
2.3.4 State and Parameter Estimates	2-31
2.4 Practical Performance Enhancements for Multiple Model Adaptive Estimation	2-33
2.4.1 Kalman Filter Tuning	2-34
2.4.2 Harmonically Balanced Kalman Filters	2-34
2.4.3 Scalar Residual Monitoring	2-35
2.4.4 β Dominance Compensation	2-36
2.4.5 Scalar Penalty Modification	2-37
2.4.6 Lower Bounding Conditional Probabilities . .	2-37
2.4.7 Markov Process Modeling of Hypothesis Condi- tional Probabilities	2-38
2.4.8 Hypothesis Swapping	2-40
2.4.9 Probability Smoothing	2-40
2.4.10 Increased Residual Propagation	2-41
2.4.11 Dithering	2-41
2.4.12 Filter Pruning	2-42
2.4.13 Filter Restart	2-43
2.4.14 Maximum Entropy with Identity Covariance .	2-43
2.4.15 Pseudo-Residuals	2-44
2.4.16 Generalized Residuals	2-44
2.5 A Virtual Filter Bank Using Only a Single Filter . . .	2-45
2.5.1 An Example: Different Input Distributor Matri- ces	2-47
2.5.2 Equivalent Residuals	2-48
2.5.3 Computational Savings	2-48

	Page
2.5.4 Comment	2-48
2.6 Moving Bank Multiple Model Adaptive Estimation Fundamentals	2-49
2.6.1 Moving-Bank versus Fixed-Bank Multiple Model Adaptive Estimation	2-50
2.6.2 A Short Glossary of Bank Manipulation Terms	2-51
2.6.3 Logic Rules for Moving the Bank	2-52
2.6.4 Hypothesis Testing	2-57
2.7 Advanced Moving-Bank Multiple Model Adaptive Estimation Structures and Techniques	2-61
2.7.1 Hierarchical Structured Filter Bank	2-61
2.7.2 Filter Spawning	2-62
2.7.3 Optimal Parameter Discretization	2-62
2.7.4 Inter-Residual Distance Feedback	2-64
2.7.5 Probability-Based Discretization Method . . .	2-67
2.7.6 Residual Correlation Kalman Filter Bank . . .	2-67
2.7.7 Modified Multiple Model Adaptive Estimation	2-69
2.8 A Final Note	2-70
III. Infinite-Dimensional Sampled-Data Kalman Filter	3-1
3.1 Introduction	3-1
3.2 Preliminary Concepts	3-2
3.3 Linear Infinite-Dimensional Minimum Variance Unbiased Estimator	3-41
3.4 Dynamics Model	3-71
3.5 Infinite-Dimensional Sampled-Data Kalman Filter . . .	3-82
3.6 Generalized Infinite-Dimensional Multiple Model Adaptive Estimation	3-92
3.6.1 Elemental Filters	3-93

	Page
3.6.2 State and Parameter Estimates	3-95
3.7 Summary	3-97
IV. An Example: The Stochastic Heat Equation	4-1
4.1 Introduction	4-1
4.2 Mathematical System Model	4-2
4.2.1 Preliminary Background: The Heat Equation	4-3
4.2.2 Model Mapping	4-4
4.2.3 Discrete-Time Measurement Model	4-5
4.2.4 Continuous-Time Dynamics Model	4-6
4.2.5 Equivalent Discrete-Time Dynamics Model . .	4-8
4.2.6 Equivalent Infinite-Dimensional Discrete-Time Model	4-9
4.2.7 Essentially-Equivalent Finite-Dimensional Discrete-Time Model	4-10
4.2.8 Approximating the Infinite-Dimensional State Via Projection	4-12
4.3 Kalman Filtering Algorithm	4-16
4.3.1 The State Transition Operator Adjoint Φ^* . .	4-16
4.3.2 The Measurement Distributor Transformation Adjoint \mathbf{H}^*	4-18
4.3.3 The Residual Covariance Matrix $\mathbf{A}(t_i)$	4-19
4.3.4 The Kalman Gain Transformation K	4-24
4.3.5 The State Transition Matrix $\tilde{\Phi}(\Delta t_{i+1})$	4-26
4.3.6 The Equivalent Discrete-Time Input Distributor Matrix $\tilde{\mathbf{B}}_d(t_i)$	4-29
4.3.7 Propagate the Finite-Dimensional State Estimate	4-31
4.3.8 The Equivalent Discrete-Time Dynamics Noise Covariance Matrix $\tilde{\mathbf{Q}}_d(t_i)$	4-31

	Page
4.3.9 First and Second Order Statistical Moments for the State Coefficients	4-34
4.3.10 The Propagation Error Covariance Matrix $\tilde{\mathbf{P}}(t_{i+1}^-)$	4-36
4.3.11 The Measurement Distributor Matrix $\tilde{\mathbf{H}}$. . .	4-39
4.3.12 The Residual Covariance Matrix $\tilde{\mathbf{A}}(t_i)$	4-42
4.3.13 The Kalman Gain Transformation Matrix $\tilde{\mathbf{K}}$.	4-44
4.3.14 The Updated State Estimate $\hat{x}(t_i^+)$	4-46
4.3.15 The Updated Error Covariance Operator $P(t_i^+)$	4-47
4.3.16 Summary	4-52
4.4 Multiple Model Adaptive Estimation	4-55
4.5 Summary	4-57
V. Simulation and Results	5-1
5.1 Introduction	5-1
5.2 Simulation 1	5-12
5.3 Simulation 2	5-27
5.4 Simulation 3	5-72
5.5 Simulation 4	5-104
5.5.1 Simulation 4, Case 1 (Increasing R_{true})	5-105
5.5.2 Simulation 4, Case 2 (Decreasing R_{true})	5-122
5.6 Simulation 5	5-138
5.7 Simulation 6	5-155
5.8 Summary	5-186
VI. Conclusions	6-1
6.1 Introduction	6-1
6.2 Contributions	6-1
6.3 Recommendations	6-4

	Page
Bibliography	BIB-1
Vita	VITA-1

List of Figures

Figure		Page
1.1.	Mapping the Infinite-Dimensional Continuous-Time Model to Finite-Dimensional Discrete-Time Models.	1-4
1.2.	First Sextant Sighting.	1-9
1.3.	Second Sextant Sighting.	1-10
1.4.	First and Second Sextant Sightings Combined.	1-13
1.5.	Multiple Model Adaptive Estimation: Two Kalman Filters in Parallel	1-16
2.1.	Multiple Model Adaptive Estimation	2-8
2.2.	Surround the true parameter.	2-26
2.3.	Coarse and fine discretization of filter bank.	2-27
3.1.	Boxology of a Transformation or Mapping.	3-3
3.2.	Boxology of a Linear Transformation.	3-4
3.3.	An Illustration of the Projection Theorem.	3-22
3.4.	Boxology of a Random Vector.	3-23
3.5.	Boxology of a Conditional Expectation.	3-36
3.6.	Boxology of a Measurement Model.	3-49
3.7.	Boxology of the LIMVUE for CSO.	3-50
3.8.	Boxology of a Generalized Linear Measurement Model.	3-57
3.9.	Boxology of the Linear Infinite-Dimensional Minimum Variance Unbiased Estimator.	3-63
3.10.	Boxology of a Generalized Linear Stochastic Measurement Model.	3-67
3.11.	Boxology of the Stochastic LIMVUE.	3-69
3.12.	Boxology of a Discrete-Time Dynamics Model.	3-75

Figure		Page
3.13.	Boxology of the ISKF.	3-88
4.1.	Mapping the Infinite-Dimensional Continuous-Time Model to an Essentially-Equivalent Finite-Dimensional Discrete-Time Model.	4-5
4.2.	Boxology of the Random Measurement Vector.	4-7
5.1.	Simulation 1: Filter bank composition experiment for $N = 30$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 0.1$	5-15
5.2.	Simulation 1: Filter bank composition experiment for $N = 30$, $Q_{\text{true}} = 100$, $R_{\text{true}} = 1$	5-16
5.3.	Simulation 1: Filter bank composition experiment for $N = 30$, $Q_{\text{true}} = 50$, $R_{\text{true}} = 1$	5-17
5.4.	Simulation 1: Filter bank composition experiment for $N = 30$, $Q_{\text{true}} = 20$, $R_{\text{true}} = 1$	5-18
5.5.	Simulation 1: Filter bank composition experiment for $N = 30$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 1$	5-19
5.6.	Simulation 1: Filter bank composition experiment for $N = 40$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 1$	5-20
5.7.	Simulation 1: Filter bank composition experiment for $N = 50$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 1$	5-21
5.8.	Simulation 1: Filter bank composition experiment for $N = 60$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 1$	5-22
5.9.	Simulation 1: Filter bank composition experiment for $N = 70$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 1$	5-23
5.10.	Simulation 1: Filter bank composition experiment for $N = 30$, $Q_{\text{true}} = 100$, $R_{\text{true}} = 1$, $R_{\text{filter}} = 10$	5-24
5.11.	Simulation 1: Filter bank composition experiment for $N = 30$, $Q_{\text{true}} = 100$, $R_{\text{true}} = 10$, $R_{\text{filter}} = 1$	5-25
5.12.	Simulation 2: System dynamics noise covariance.	5-29
5.13.	Simulation 2: Initial state covariance experiment — hypothesis conditional probability flow.	5-30

Figure		Page
5.14.	Simulation 2, Case 1 (P_0^{poor}): Elemental Filter 1.	5-36
5.15.	Simulation 2, Case 1 (P_0^{poor}): Elemental Filter 2.	5-38
5.16.	Simulation 2, Case 1 (P_0^{poor}): Elemental Filter 3.	5-40
5.17.	Simulation 2, Case 1 (P_0^{poor}): Blended Filter.	5-42
5.18.	Simulation 2, Case 2 (P_0^{good}): Elemental Filter 1.	5-45
5.19.	Simulation 2, Case 2 (P_0^{good}): Elemental Filter 2.	5-47
5.20.	Simulation 2, Case 2 (P_0^{good}): Elemental Filter 3.	5-49
5.21.	Simulation 2, Case 2 (P_0^{good}): Blended Filter.	5-51
5.22.	Simulation 2, Case 3 (P_0^{poor}): Elemental Filter 1.	5-54
5.23.	Simulation 2, Case 3 (P_0^{poor}): Elemental Filter 2.	5-56
5.24.	Simulation 2, Case 3 (P_0^{poor}): Elemental Filter 3.	5-58
5.25.	Simulation 2, Case 3 (P_0^{poor}): Blended Filter.	5-60
5.26.	Simulation 2, Case 4 (P_0^{good}): Elemental Filter 1.	5-64
5.27.	Simulation 2, Case 4 (P_0^{good}): Elemental Filter 2.	5-66
5.28.	Simulation 2, Case 4 (P_0^{good}): Elemental Filter 3.	5-68
5.29.	Simulation 2, Case 4 (P_0^{good}): Blended Filter.	5-70
5.30.	Simulation 3: Measurement-corruption noise covariance. . . .	5-73
5.31.	Simulation 3: Hypothesis conditional probability flow. . . .	5-75
5.32.	Simulation 3, Case 1: Elemental Filter 1.	5-78
5.33.	Simulation 3, Case 1: Elemental Filter 2.	5-80
5.34.	Simulation 3, Case 1: Elemental Filter 3.	5-82
5.35.	Simulation 3, Case 1: Elemental Filter 4.	5-84
5.36.	Simulation 3, Case 1: Elemental Filter 5.	5-86
5.37.	Simulation 3, Case 1: Blended Filter.	5-88
5.38.	Simulation 3, Case 2: Elemental Filter 1.	5-92
5.39.	Simulation 3, Case 2: Elemental Filter 2.	5-94
5.40.	Simulation 3, Case 2: Elemental Filter 3.	5-96

Figure		Page
5.41.	Simulation 3, Case 2: Elemental Filter 4.	5-98
5.42.	Simulation 3, Case 2: Elemental Filter 5.	5-100
5.43.	Simulation 3, Case 2: Blended Filter.	5-102
5.44.	Simulation 4, Case 1 (increasing R_{true}): Hypothesis conditional probability.	5-106
5.45.	Simulation 4, Case 1 (increasing R_{true}): Likelihood quotient. .	5-108
5.46.	Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 1. .	5-110
5.47.	Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 2. .	5-112
5.48.	Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 3. .	5-114
5.49.	Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 4. .	5-116
5.50.	Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 5. .	5-118
5.51.	Simulation 4, Case 1 (increasing R_{true}): Blended Filter. . . .	5-120
5.52.	Simulation 4, Case 2 (decreasing R_{true}): Hypothesis conditional probability.	5-122
5.53.	Simulation 4, Case 1 (decreasing R_{true}): Likelihood quotient.	5-124
5.54.	Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 1. .	5-126
5.55.	Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 2. .	5-128
5.56.	Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 3. .	5-130
5.57.	Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 4. .	5-132
5.58.	Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 5. .	5-134
5.59.	Simulation 4, Case 2 (decreasing R_{true}): Blended Filter. . . .	5-136
5.60.	Simulation 5: Hypothesis conditional probability.	5-139
5.61.	Simulation 5: Abruptly changing measurement-corruption noise covariance.	5-139
5.62.	Simulation 5: Elemental Filter 1.	5-143
5.63.	Simulation 5: Elemental Filter 2.	5-145
5.64.	Simulation 5: Elemental Filter 3.	5-147
5.65.	Simulation 5: Elemental Filter 4.	5-149

Figure		Page
5.66.	Simulation 5: Elemental Filter 5.	5-151
5.67.	Simulation 5: Blended Filter.	5-153
5.68.	Simulation 6: Rod heating profile.	5-157
5.69.	Simulation 6: Hypothesis conditional probability.	5-158
5.70.	Simulation 6 (maximum excitation): Elemental Filter 1. . . .	5-162
5.71.	Simulation 6 (maximum excitation): Elemental Filter 2. . . .	5-164
5.72.	Simulation 6 (maximum excitation): Elemental Filter 3. . . .	5-166
5.73.	Simulation 6 (maximum excitation): Elemental Filter 4. . . .	5-168
5.74.	Simulation 6 (maximum excitation): Elemental Filter 5. . . .	5-170
5.75.	Simulation 6 (maximum excitation): Blended Filter.	5-172
5.76.	Simulation 6 (minimum excitation): Elemental Filter 1. . . .	5-174
5.77.	Simulation 6 (minimum excitation): Elemental Filter 2. . . .	5-176
5.78.	Simulation 6 (minimum excitation): Elemental Filter 3. . . .	5-178
5.79.	Simulation 6 (minimum excitation): Elemental Filter 4. . . .	5-180
5.80.	Simulation 6 (minimum excitation): Elemental Filter 5. . . .	5-182
5.81.	Simulation 6 (minimum excitation): Blended Filter.	5-184

List of Tables

Table		Page
3.1.	Quartet of Kalman Filters	3-83
5.1.	Truth Parameters for the simulations	5-3
5.2.	Arrangement for Plots (a) through (h) for the k th Elemental Filter	5-5
5.3.	Simulation 1: Four key parameters for the filter bank composition experiment	5-13
5.4.	Simulation 1: Elemental filter parameters for the filter bank composition experiment.	5-14
5.5.	Dynamics noise strengths for each of the elemental filters shown in the 16 plots of Figure 5.1, where $Q_{\text{true}} = 10$ and $R_{\text{true}} = 0.1$	5-14
5.6.	Simulation 2: Elemental filter parameters for the initial state error covariance experiment.	5-28
5.7.	Simulation 3: Elemental Filter Parameters	5-73
5.8.	Simulation 4: Elemental Filter Parameters	5-104
5.9.	Simulation 4, Case 1 (increasing R_{true}): Best hypothesis.	5-107
5.10.	Simulation 4, Case 2 (decreasing R_{true}): Best hypothesis.	5-123
5.11.	Simulation 5: Elemental Filter Parameters	5-138
5.12.	Simulation 5: Expected likelihood quotients	5-140
5.13.	Simulation 6: Truth Parameters	5-155
5.14.	Simulation 6: Elemental Filter Parameters	5-155

List of Symbols

Symbol		Page
\mathbb{A}	parameter space	2-21
\mathbb{A}_j	parameter space for j th parameter	2-21
\mathbf{A}_k	k th filter-computed residual covariance matrix	2-8
$\mathbf{A}(t_i)$	filter-computed residual covariance at time t_i	2-17
$\mathbf{A}(t_i^+)$	post-fit residual covariance matrix at time t_i	2-45
$\tilde{\mathbf{A}}(t_i)$	matrix representation of $\mathbf{A}(t_i)$	4-44
$\mathbf{A}_{\text{true}}(t_i)$	true filter-computed residual covariance at time t_i	5-10
\mathbf{A}_k^{ss}	k th steady state filter-computed residual covariance	5-11
$\mathbf{A}_{\text{true}}^{\text{ss}}$	true steady state filter-computed residual covariance	5-11
\mathbf{a}_k	k th system mode parameter vector	2-8
$\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_K$	set of K parameter vectors	2-7
$\mathbf{a}(t_i)$	stochastic system mode at time t_i	2-18
$\dot{\hat{\mathbf{a}}}(t_i)$	parameter estimate velocity at time t_i	2-54
$\hat{\mathbf{a}}$	parameter vector estimate	2-8
$\hat{\mathbf{a}}_{\text{MAP-MMAE}}(t_i^+)$	MAP parameter estimate from MMAE at time t_i	2-33
$\mathbf{B}(t)$	input distributor matrix at time t	2-9
$\mathbf{B}_d(t_i)$	discrete-time input distributor matrix at time t_i	2-11
$\tilde{\mathbf{B}}_d(t_i)$	matrix representation of $\mathbf{B}_d(t_i)$	4-30
B	bounded linear transformation	3-6
B^*	adjoint of BLT B	3-12
$\mathfrak{B}(\mathbb{A})$	Borel sets of set \mathbb{A}	3-24
\mathbb{B}_N	orthonormal basis on a subspace	4-14
\mathbb{B}_∞	orthonormal basis	4-13
$\mathcal{BLO}(\mathbb{X})$	space of bounded linear operators	3-6
$\mathcal{BLT}(\mathbb{X}, \mathbb{Y})$	space of bounded linear transformations	3-6

Symbol		Page
$\mathbf{b}(t)$	Brownian motion (or Wiener) process	2-10
$C([a, b])$	continuous real-valued functions defined on an interval	3-10
$C(\mathbf{e})$	cost function	3-45
$C_{\mathbf{a}}$	cost functional	2-63
\mathbb{C}	set of complex numbers	3-3
\mathbf{c}	estimator bias	3-48
D	MMAE decision block	2-8
$\mathcal{D}(L)$	domain of the operator L	3-15
E	expectation operator	3-25
$E_{\mathbf{y}}(\mathbf{y})$	expectation with respect to \mathbf{y}	3-25
$E_{\mathbf{x}, \mathbf{y}}$	joint expectation	3-26
\mathcal{E}	conditional expectation operator	3-34
$\mathcal{E}(\mathbf{x} \mathcal{S})$	conditional expectation	3-34
\mathbf{e}	estimator error	3-45
$\mathbf{F}(t)$	system dynamics matrix at time t	2-9
$\mathbf{F}_j(t)$	system dynamics matrix at time t	1-2
F	second-order partial differential operator	4-7
F_1, F_2, \dots, F_K	set of K elemental filters	2-7
$\mathfrak{F}(\cdot, \cdot)$	linear space of functions	3-23
\mathbb{F}	scalar field	3-8
\mathcal{F}	σ -field	3-25
\mathcal{F}_1	filtering process	1-3
\mathcal{F}_{opt}	optimal filtering process	1-3
$f_{\mathbf{x}(t_0)}(\boldsymbol{\xi})$	probability density function	2-16
$f_{\mathbf{x}(t_i) \mathbf{z}(t_i)}(\boldsymbol{\xi} \mathbf{Z}_i)$	conditional probability density function	2-17
$\mathbf{G}(t)$	noise distributor matrix at time t	2-9
$\mathbf{G}_d(t_i)$	discrete-time noise distributor matrix at time t_i . . .	2-11

Symbol		Page
\mathcal{G}	σ -algebra	3-24
\mathbf{H}	measurement distributor transformation	3-56
$\mathbf{H}(t_i)$	measurement distributor matrix at time t_i	2-9
$\tilde{\mathbf{H}}$	matrix representation of \mathbf{H}	4-41
\mathbf{I}	identity matrix	2-11
$J_{jk}(t_i)$	inter-residual difference measure at time t_i	2-65
$\mathfrak{J}(\mathbf{c})$	cost function	3-51
$\mathbf{K}(t_i)$	Kalman gain at time t_i	2-17
$\tilde{\mathbf{K}}(t_i)$	matrix representation of $\mathbf{K}(t_i)$	4-46
K	number of elemental filters in an MMAE filter bank	2-7
k^*	filter with max. hypothesis conditional probability	2-33
L	linear transformation	3-3
$L_k(t_i)$	k th likelihood quotient at time t_i	2-30
$\mathbf{L}_k(t_i)$	k th random likelihood quotient at time t_i	5-10
$\mathbb{L}^2([0, 1], \mathbb{R})$	space of Lebesgue measurable \mathbb{L}^2 functions for $0 \leq \rho \leq 1$	4-6
$\mathbb{L}_{[a,b]}^p$	space of Lebesgue measurable functions over interval	3-7
$\mathcal{LO}(\mathbb{Y})$	space of linear operators	3-4
$\mathcal{LT}(\mathbb{X}, \mathbb{Y})$	space of linear transformations	3-4
$l_k(t_i)$	proximity of the k th filter at time t_i	2-64
\mathbf{M}^*	conjugate transpose of matrix \mathbf{M}	3-15
M	integer number of sensors, sections, or measurements	4-5
m, n	indices (integers)	3-7
N	integer	3-7
$\mathfrak{N}[\cdot; \cdot, \cdot]$	Gaussian (normal) probability distribution	2-8
\mathbb{N}	natural numbers, positive integers	3-7
$\mathcal{NT}(\mathbb{X}, \mathbb{Y})$	nuclear transformation from \mathbb{X} to \mathbb{Y}	3-18
$\mathcal{O}(\mathbb{Y})$	space of operators	3-3

Symbol		Page
\mathbf{P}_0	initial state covariance estimate	2-13
$\mathbf{P}(t_i^-)$	propagated state error covariance at time t_i	2-16
$\mathbf{P}(t_i^+)$	measurement-updated state error covariance at time t_i	2-17
$\mathbf{P}_k(t_i^+)$	k th state error covariance at time t_i	2-32
$\mathbf{P}_{\mathbf{a},\text{MMAE}}(t_i^+)$	conditional covariance of $\mathbf{a}(t_i)$ from MMAE	2-33
$\mathbf{P}_{\text{MMAE}}(t_i^+)$	conditional covariance of $\mathbf{x}(t_i)$ from MMAE	2-32
$\tilde{\mathbf{P}}(t_i^-)$	matrix representation of $\mathbf{P}(t_i^-)$	4-36
$\tilde{\mathbf{P}}(t_i^+)$	matrix representation of $\mathbf{P}(t_i^+)$	4-36
$\tilde{\mathbf{P}}_k(t_i^+)$	k th state coefficient error covariance at time t_i	4-56
$\tilde{\mathbf{P}}_{\text{MMAE}}(t_i^+)$	matrix representation of $\mathbf{P}_{\text{MMAE}}(t_i^+)$	4-56
P	probability measure	3-25
$P_{\mathbf{y}}$	probability measure induced by \mathbf{y}	3-25
$P_{\mathbf{x},\mathbf{y}}$	joint probability measure	3-26
\mathcal{P}	projection operator	4-14
$\mathbf{p}(t_i)$	hypothesis conditional probability vector	2-39
p_k	hypothesis conditional probability for k th elemental filter	2-8
$\text{pr}\{\cdot\}$	probability	1-18
$\mathbf{Q}(t)$	continuous-time dynamics noise strength at time t	2-12
$\mathbf{Q}_d(t_i)$	discrete-time dynamics noise covariance at time t_i	2-12
$\tilde{\mathbf{Q}}_d(t_i)$	matrix representation of $\mathbf{Q}_d(t_i)$	4-34
$\mathbf{R}(t_i)$	measurement noise covariance at time t_i	2-12
\mathbf{R}_{true}	true measurement-corruption noise covariance	5-11
R	measurement-corruption noise covariance eigenvalue	5-4
R_{median}	median measurement-corruption noise covariance	5-104
\mathbb{R}	set of real numbers	3-3
\mathbb{R}^N	Euclidean space of real-valued N -vectors	3-9
$\mathcal{R}(\mathbf{y})$	range of \mathbf{y}	3-35

Symbol		Page
$\mathbf{r}(t_i)$	measurement residual at time t_i	2-17
$\mathbf{r}(t_i^+)$	post-fit residual at time t_i	2-44
$\tilde{\mathbf{r}}(t_i)$	approximate measurement residual at time t_i	4-46
$\mathbf{r}^*(t_i)$	generalized residual at time t_i	2-45
\mathbf{r}_k	k th measurement residual	2-8
$\mathbf{r}_{jk}(t_i)$	inter-residual difference at time t_i	2-65
$\mathbf{r}_k(t_i)$	k th random measurement residual at time t_i	5-10
\mathcal{S}	sub σ -field	3-34
\mathcal{S}_1	spatial process	1-4
\mathcal{S}_{opt}	optimal spatial process	1-4
\mathbf{s}_k	new information	2-65
$\mathbf{T}(t_i, t_{i-1})$	transition probability matrix	2-39
\mathbf{T}	matrix (or vector) transpose operator	3-9
T	transformation	3-3
$T_{mn}(t_i, t_{i-1})$	transition probability element	2-39
T_L	likelihood quotient threshold level	2-52
T_{χ^2}	chi-squared test threshold	2-59
\mathcal{T}_1	temporal process	1-4
\mathcal{T}_{opt}	optimal temporal process	1-4
$\mathcal{T}(\mathbb{X}, \mathbb{Y})$	space of transformations	3-3
t	time	1-2
t_i^-	time t_i prior to measurement update	2-16
t_i^+	time t_i after measurement update	2-16
$\text{tr}\{\cdot\}$	trace operator	3-19
\mathbf{u}	control input	2-7
$\mathbf{u}(t)$	control vector at time t	2-9
$\mathbf{u}(t, \rho)$	input at time t and position ρ	1-2

Symbol		Page
$\mathbf{u}(t_i)$	discrete-time control vector at time t_i	2-11
$\tilde{u}(t_i)$	approximate input function at time t_i	4-12
\mathbf{v}	white Gaussian measurement-corruption noise vector	3-56
$\mathbf{v}(t_i)$	white Gaussian measurement noise vector at time t_i .	2-9
\mathbf{W}_a	weighting matrix	2-63
$\mathbf{w}(t)$	white Gaussian noise process vector at time t	2-9
$\mathbf{w}_d(t_i)$	discrete-time white Gaussian noise vector at time t_i .	2-11
$\tilde{\mathbf{w}}_d(t_i)$	approximate discrete-time dynamics noise at time t_i .	4-12
\mathbb{X}	vector space	3-3
\mathbb{X}^*	dual space of Banach space \mathbb{X}	3-17
$\mathbb{X} \oplus \mathbb{Y}$	direct sum of \mathbb{X} and \mathbb{Y}	3-6
$\mathbb{X} \times \mathbb{Y}$	Cartesian product of \mathbb{X} and \mathbb{Y}	3-6
(\mathbb{X}, d)	metric space	3-4
$(\mathbb{X}, \mathcal{G})$	measurable space	3-24
$(\mathbb{X}, \mathcal{G}, \mu)$	measure space	3-24
\mathbf{x}	state vector	3-42
$\hat{\mathbf{x}}$	estimator	3-42
$\mathbf{x}(\cdot, \cdot)$	stochastic (state) process	3-37
$\mathbf{x}(t, \cdot)$	random (state) vector	3-37
$\mathbf{x}(t)$	state vector at time t	2-9
$\dot{\mathbf{x}}(t)$	time derivative of state, \mathbf{x}	2-9
$\mathbf{x}(\omega)$	and $\mathbf{x}(\cdot, \omega)$ — a sample of the stochastic process \mathbf{x} . .	3-37
$\mathbf{x}(t)$	and $\mathbf{x}(t, \omega)$ — a realization of the random vector $\mathbf{x}(t)$	3-37
$\mathbf{x}(t, \rho)$	state at time t and position ρ	1-2
$\mathbf{x}(\omega) = x$	a realization of stochastic process \mathbf{x}	1-24
$\tilde{\mathbf{x}}(t_i)$	approximate state function at time t_i	4-12
$\hat{\mathbf{x}}$	state estimate	2-8

Symbol		Page
$\hat{\mathbf{x}}_0$	initial state estimate	2-13
$\hat{\mathbf{x}}_k$	k th elemental filter state estimate	2-7
$\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_K$	set of K state estimates	2-8
$\hat{\mathbf{x}}(t_i^-)$	propagated state estimate at time t_i	2-16
$\hat{\mathbf{x}}(t_i^+)$	measurement-updated state estimate at time t_i	2-17
$\hat{\mathbf{x}}_{\text{MAP-MMAE}}(t_i^+)$	MAP state estimate from MMAE at time t_i	2-33
x	scalar state variable: temperature	4-3
$x(t_i, \rho)$	state function at time t_i and position ρ	4-12
$\tilde{x}(t_i, \rho)$	approximate state function at time t_i and position ρ .	4-14
\mathbb{Y}	vector space	3-3
$(\mathbb{Y}, \ \cdot\)$	normed linear space	3-5
$(\mathbb{Y}, \langle \cdot, \cdot \rangle)$	inner product space	3-8
$\mathbf{Z}(t_i)$	stochastic measurement history	2-14
\mathbf{Z}_i	measurement history realization	2-14
\mathbb{Z}	observation space	3-23
\mathbf{z}	measurement realization	2-7
\mathbf{z}	measurement vector	3-56
$\mathbf{z}(t_i)$	measurement vector at time t_i	2-9
α	dummy variable for the system mode	2-18
$\alpha(t_i)$	state coefficient vector at time t_i	4-14
$\alpha_n(t_i)$	n th Fourier series coefficient at time t_i	4-12
$\beta_k(t_i)$	k th scale factor (of a Gaussian PDF) at time t_i	2-30
$\beta_n(\rho)$	basis element	4-12
$\beta(\rho)$	basis vector	4-14
γ	generalized residual scale factor	2-45
δ_{ij}	Kronecker delta	2-9
$\delta(t)$	Dirac delta function	2-10

Symbol		Page
$\delta\mathfrak{J}(\mathbf{K}; \mathbf{L})$	Gateaux variation of \mathfrak{J}	3-54
$\Delta\mathbf{B}_{kj}$	Delta \mathbf{B} between j th and k th filters	2-47
$\Delta\mathbf{H}_{kj}$	Delta \mathbf{H} between j th and k th filters	2-47
$\Delta\mathbf{K}_{kj}$	Delta \mathbf{K} between j th and k th filters	2-47
$\Delta\boldsymbol{\epsilon}_{jk}(t_i^+)$	Delta error between j th and k th filters	2-46
$\Delta\boldsymbol{\Phi}_{kj}$	Delta $\boldsymbol{\Phi}$ between j th and k th filters	2-47
ϵ	“small” positive real number	3-7
ζ_i	dummy variable for the measurement process at time t_i	2-30
$\eta_m(t_i^-)$	integrated error at time t_i^-	4-22
κ	thermal diffusivity constant	4-3
μ	measure	3-24
$\boldsymbol{\mu}_{\mathbf{y}}$	mean of \mathbf{y}	3-25
$\mu_{m,n}$	integrated basis	4-40
$\boldsymbol{\mu}_m$	integrated basis vector	4-40
$\Xi(\mathbf{x})$	correlation operator	3-29
$\Xi(\mathbf{x}, \mathbf{y})$	cross-correlation transformation	3-30
ξ	dummy variable for the state	2-15
ρ	position	1-2
$\Sigma(\mathbf{x})$	covariance operator	3-29
$\Sigma(\mathbf{x}, \mathbf{y})$	cross-covariance transformation	3-29
$\tilde{\Sigma}(\mathbf{x} \mathcal{S})$	conditional covariance operator	3-36
$\Sigma[\mathbf{x}(t), \mathbf{x}(\tau)]$	covariance kernel operator	3-37
Σ	MMAE summing block	2-8
τ	time delay constant	1-2
$\boldsymbol{\Phi}(t_{i+1}, t_i)$	state transition matrix from time t_i to time t_{i+1}	2-11
$\tilde{\boldsymbol{\Phi}}(\Delta t_{i+1})$	matrix representation of state transition operator	4-26
χ^2	chi-squared random variable	2-59

Symbol		Page
$\chi_{\mathbf{x}}$	characteristic functional	3-32
Ω	sample space	3-23
(Ω, \mathcal{F}, P)	probability space	3-25
ω	an element of the sample space Ω	3-23
\emptyset	empty or null set	3-24
$\frac{\partial}{\partial t}$	partial derivative with respect to time t	1-2
$\frac{\partial}{\partial \rho}$	partial derivative with respect to position ρ	1-2
$\frac{\partial^2}{\partial \rho^2}$	second partial derivative with respect to position ρ	4-3
$\frac{d}{dt}$	derivative with respect to time t	1-2
\cong	approximately equal	1-10
\triangleq	defined as	1-18
\equiv	equivalent to	1-18
$\ \cdot\ $	norm	3-5
$[a, b]$	closed real interval from a to b	3-8
$\langle \cdot, \cdot \rangle$	inner product	3-8
\perp	orthogonal vectors, sets, or spaces	3-9
$\mathbf{x} \diamond \mathbf{y}$	outer product of \mathbf{x} and \mathbf{y}	3-9
$^{\circ}\text{C}$	degrees Celsius	4-3

List of Abbreviations

Abbreviation		Page
ACP	abstract Cauchy problem	1-23
AIMMAE	approximate infinite-dimensional MMAE	5-1
BLO	bounded linear operator	3-6
BLT	bounded linear transformation	3-6
BLUE	best linear unbiased estimator	3-61
CSO	correlated states and observations	3-41
DPS	distributed parameter system	1-1
ECG	electrocardiogram	2-40
EKF	extended Kalman filter	2-15
GIMMAE	generalized infinite-dimensional MMAE	1-27
GLR	generalized likelihood ratio	2-60
GPB	generalized pseudo Bayesian	2-40
GPS	Global Positioning System	2-42
IID	independent, identically distributed	2-13
IKBF	infinite-dimensional Kalman-Bucy filter	1-1
IMM	interacting multiple model	2-6
IMVUE	infinite-dimensional MVU estimator	3-61
IRDF	inter-residual distance feedback	2-64
ISKF	infinite-dimensional sampled-data Kalman filter	1-3
LIMVUE	linear infinite-dimensional MVU estimator	3-2
LMVUE	linear MVU estimator	3-61
LPS	lumped parameter system	1-1
LS	least squares	3-41
M ³ AE	Modified MMAE	2-69
MAP	maximum <i>a posteriori</i>	2-15

Abbreviation		Page
ME/I	maximum entropy with identity covariance	2-44
MMAC	multiple model adaptive control	2-5
MMAE	multiple model adaptive estimation	1-1
MMSE	minimum mean-squared error	2-31
MSE	mean-squared error	2-15
MVU	minimum variance unbiased	2-15
ML	maximum likelihood	2-15
NT	nuclear transformation	3-18
PBDM	probability-based discretization method	2-67
PCA	periodically correlated acceleration	2-34
PDE	partial differential equation	1-2
PDF	probability density function	1-9
PSD	power spectral density	1-12
RCKFB	residual correlation Kalman filter bank	2-68
RMS	root mean square	5-8
SDE	stochastic differential equation	3-1
SPDE	stochastic partial differential equation	3-1
SPRT	sequential probability ratio test	2-60
SRDE	stochastic retarded differential equation	3-1
TDE	time-delayed differential equation	1-2
WGAN	white Gaussian additive noise	3-56

Abstract

Kalman filtering and multiple model adaptive estimation (MMAE) methods have been applied by researchers in several engineering disciplines to a multitude of problems as diverse as aircraft flight control and drug infusion monitoring. MMAE methods have been used to adapt to an uncertain noise environment and/or identify important system parameters in these problems. All of the model-based estimation (and control) problems considered in this earlier research have at their core a linear (or mildly nonlinear) model based on finite-dimensional differential (or difference) equations perturbed by random inputs (noise). However, many real-world systems are more naturally modelled using an infinite-dimensional continuous-time linear systems model, such as those most naturally modelled as partial differential equations or time-delayed differential equations. Thus, we are motivated to extend existing finite-dimensional techniques, such as the Kalman filter, to allow the engineer to apply familiar tools to a larger class of problems.

The focus of this research is (1) to extend the Kalman filtering technique to encompass infinite-dimensional continuous-time systems with sampled-data measurements and (2) to approximate the infinite-dimensional continuous-time system model descriptions with an essentially equivalent finite-dimensional discrete-time model upon which a filtering algorithm could be based.

The infinite-dimensional sampled-data Kalman filter (ISKF) is a mathematical extension of the finite-dimensional sampled-data Kalman filter. The ISKF is rigorously developed using the definition-theorem-proof format. First, we derive the linear infinite-dimensional minimum variance unbiased estimator (LIMVUE) based on a dynamics model driven by a Wiener process (Brownian motion) and based on the Classical Projection Theorem to handle the state estimator's measurement update cycle. Then we create an equivalent discrete-time model description based on

the problem's natural continuous-time model to provide a means to propagate the state estimator between measurement updates.

Next, the algorithm to create an essentially equivalent finite-dimensional discrete-time model from an infinite-dimensional continuous-time model is constructed by combining an existing technique for producing an equivalent discrete-time model for a finite-dimensional system and a novel Galerkin-like technique for a stochastic differential equation that completely captures the important qualities of the original infinite-dimensional description.

An extended example featuring these new tools is presented for a stochastic partial differential equation. Specifically, the temperature profile along a slender rod is estimated using a Kalman filter for the case of a one-dimensional stochastic heat equation with Neumann boundary conditions. Additionally, the MMAE with a bank of Kalman filters is used to estimate the heat profile in the face of an unknown noise environment (zero-mean white Gaussian noises with uncertain covariances in the dynamics and/or measurement models) and to perform system identification (to determine the thermal diffusivity constant) in the face of an unknown noise environment.

SAMPLED-DATA KALMAN FILTERING AND MULTIPLE MODEL ADAPTIVE ESTIMATION FOR INFINITE-DIMENSIONAL CONTINUOUS-TIME SYSTEMS

I. Introduction

In the 1960's, Kalman, Bucy, and Falb [95, 96, 51] devised what we shall call the (sampled-data measurement) Kalman filter, the (continuous-time measurement) Kalman-Bucy filter, and the infinite-dimensional (continuous-time measurement) Kalman-Bucy filter (IKBF), respectively. Shortly after the finite-dimensional filters were put forth, Magill [125] introduced a nonlinear technique to address the case of uncertain model parameters using a bank of Kalman filters. This nonlinear technique is now known as multiple model adaptive estimation (MMAE). Kalman filtering and multiple model methods used to adapt to an uncertain noise environment and/or identify important system parameters have been applied to dozens of problems in many engineering disciplines; several examples of these applications and the MMAE theory in general are presented to the reader in Chapter II. All of the model-based estimation (and control) problems considered in the research discussed in Chapter II have at their core a linear (or mildly nonlinear) model based on a system of finite-dimensional differential (or difference) equations perturbed by random inputs (noise). In contrast to finite-dimensional or lumped-parameter system (LPS) theory in which the spatial behavior of the system is concentrated at a single point in space, the field of infinite-dimensional or distributed-parameter system (DPS) theory is concerned with the dynamic behavior of processes distributed in space as well as evolving in time. While continuous-time systems of the lumped parameter variety are adequately modeled with systems of ordinary differential equations, many real-

world systems are more naturally modeled using an infinite-dimensional DPS model such as a partial differential equation (PDE), like the heat equation given by

$$\frac{\partial}{\partial t} \mathbf{x}(t, \rho) = \frac{\partial}{\partial \rho} \mathbf{x}(t, \rho) + \mathbf{u}(t, \rho) \quad (1.1)$$

where the temperature, $\mathbf{x}(t, \rho)$, is called the state and $\mathbf{u}(t, \rho)$ is some control input, or by a time-delayed differential equation (TDE), such as

$$\frac{d}{dt} \mathbf{x}(t) = \mathbf{F}_1(t) \mathbf{x}(t) + \mathbf{F}_2(t) \mathbf{x}(t - \tau) + \mathbf{u}_1(t) + \mathbf{u}_2(t - \tau) \quad (1.2)$$

where the state and control input are partitioned into current time and portions delayed by amount τ and $\mathbf{F}_j(t)$ for $j = 1, 2$ represents the system dynamics. In Chapter III, we shall generalize these equations for a stochastic state with random additive disturbances.

Two good places for an engineer to begin an investigation of DPSs are the compilations edited by Ray and Lainiotis [163] in 1978 that consists of broad chapter-long surveys written by DPS experts that fully cover the entire scope of DPS theory at that time, and the two-volume collection of benchmark papers put together by Stavroulakis [184, 185] a few years later. The notes by Curtain and Pritchard [38] presents a solid continuous-time system mathematical foundation for infinite-dimensional linear system theory. Note that, generally speaking, there are two main camps of researchers at work in this field: those concerned with the practical implementation of a solution for an application most often refer to the field as DPS theory, while those more interested in the theoretical or mathematical foundations talk of infinite-dimensional linear systems theory. However, the terms are most often treated as synonyms in the literature, as they are in this research.

We are interested in both areas of research, hence the focus of this research is twofold. On the mathematical foundations side, we begin by extending the Kalman filtering technique to encompass infinite-dimensional sampled-data mea-

surement systems: that is, we will derive, in Chapter III, the infinite-dimensional sampled-data Kalman filter (ISKF), thus completing the mathematical quartet of filters begun over forty years ago. Then we give a method for mapping an infinite-dimensional continuous-time model to an equivalent infinite-dimensional discrete-time model, so that the new ISKF can be used for both continuous-time and discrete-time models. On the practical side, we use a subspace spectral method, in the spirit of the Galerkin technique [62] for stochastic differential equations, to create an essentially-equivalent finite-dimensional discrete-time model from the equivalent infinite-dimensional discrete-time model; this approximation completely captures the important qualities of the original infinite-dimensional description. Thus, we have crafted a new method for transforming a DPS problem into an LPS problem that can be solved using existing tools and techniques.

1.1 Overview

The primary purpose of this research is to extend the applicability of Kalman filtering and MMAE to problems well-modeled using infinite-dimensional continuous-time linear systems with sampled-data measurements¹. In this research, we derive the ISKF algorithm — this is accomplished in Chapter III. The ISKF can be applied to a large class of DPS problems modeled by an infinite-dimensional stochastic differential equation.

In Figure 1.1 we have captured the primary solution paths taken by other researchers in the top two paths and this research in the bottom path to solve infinite-dimensional problems using finite-dimensional tools ($\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_{\text{opt}}$). From left to right, the path begins with a projection of the “truth” onto an infinite-dimensional continuous-time system and discrete-time (sampled-data) measurement model. The top two paths conceptually represent existing suboptimal methods used to map the infinite-dimensional continuous-time model to a finite-dimensional discrete-time

¹Infinite-dimensional *discrete-time* linear systems with discrete-time measurements are also covered.

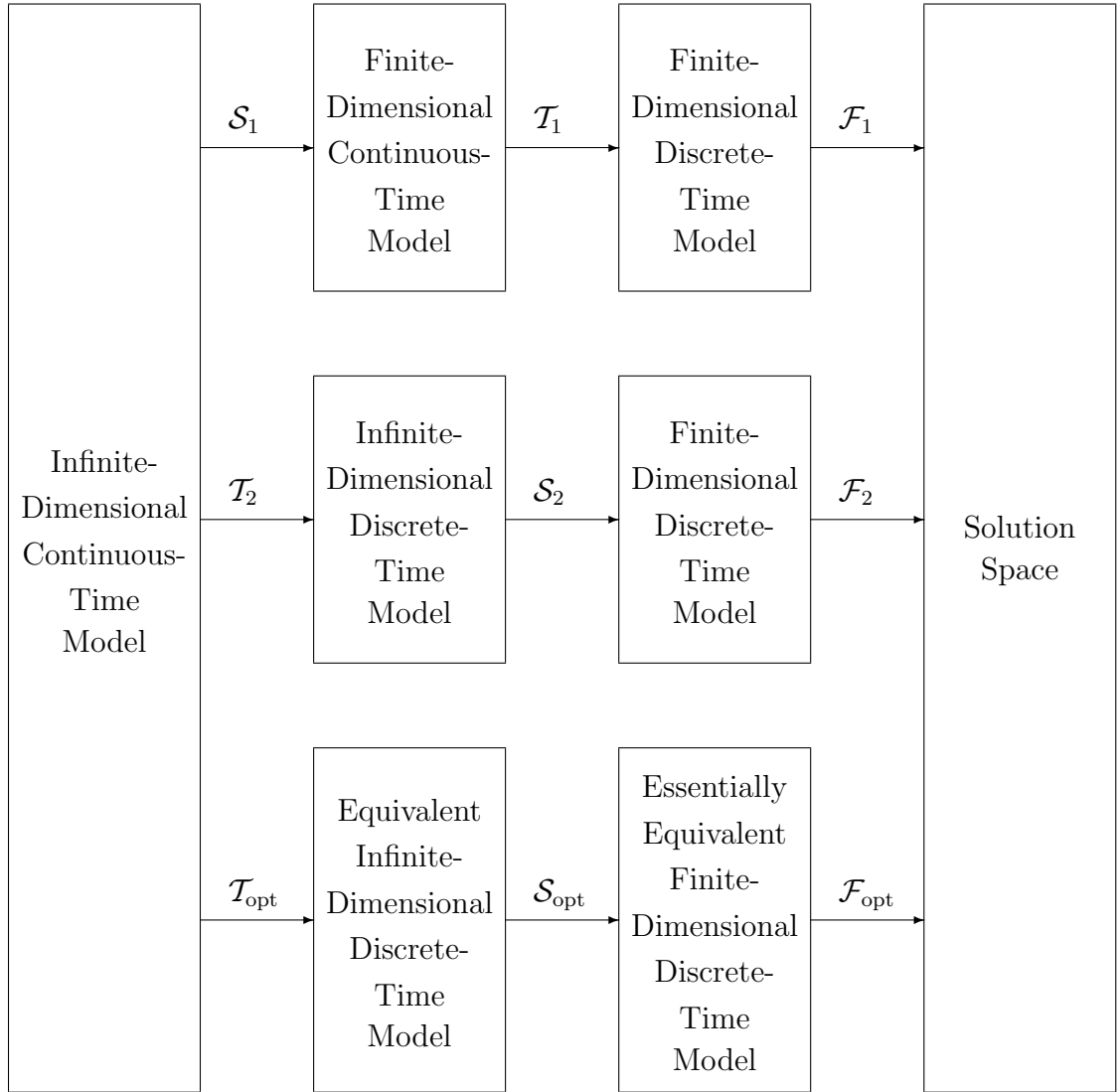


Figure 1.1 Mapping the Infinite-Dimensional Continuous-Time Model to Finite-Dimensional Discrete-Time Models.

model so that a digital filtering algorithm can be used to estimate the state and/or parameter of interest associated with the approximate model. The top path could represent an infinite-dimensional system modeled by a partial differential equation with stochastic (and perhaps deterministic) inputs. Then, \mathcal{S}_1 represents the process of approximating the *spatial* (and thus the notation \mathcal{S}_1) partial derivatives, thus reducing the dimension of the model to some finite number. Finite elements [89, 30]

and finite differences [44, 111] are common approximation methods. A discrete-time model (that is perhaps equivalent in the same sense as in our path as explained below) could be found next, and thus \mathcal{T}_1 could be \mathcal{T}_{opt} or any other *ad hoc* technique. The middle path could be used to demonstrate the time-delay problem, modeled by a stochastic retarded differential equation. The measurement time delay issue could be approximated by \mathcal{T}_2 . Then \mathcal{S}_2 would be an identity operation as it is not needed for this problem.

The bottom path is fully developed in Chapter IV where we demonstrate our technique for creating an *equivalent* infinite-dimensional *discrete-time* model (denoted by \mathcal{T}_{opt} on Figure 1.1) from the original infinite-dimensional continuous-time model. The model is termed equivalent² because the (infinite-dimensional) state³ is identical in the continuous-time and discrete-time models at an arbitrary time instant. Thus an equivalent discrete-time model properly characterizes the continuous-time dynamics model of the system. Next, we demonstrate how to create an *essentially-equivalent* finite-dimensional discrete-time model by *projecting* the infinite-dimensional model onto a finite-dimensional subspace — denoted by \mathcal{S}_{opt} . We need a finite-dimensional subspace so that we can use a digital computer to implement an algorithm. How the projection is undertaken is very important because the remaining dimensions of the infinite-dimensional model should be those “directions” of the vector space which are dominated by the noise inputs (i.e., the uncertainties in

²When we have a continuous-time dynamics model, we follow Maybeck [129] and create an equivalent discrete-time model prior to designing a digital filter to process the data optimally. A less desirable method would be to design a continuous-time filter matched to the continuous-time model and then discretize the filter to allow computation on a digital computer. This second method involves numerically solving a Riccati equation — this should be avoided if at all possible because these solutions are often unstable (despite theoretical solution stability); hence the solution might not converge. Moreover, the discretized version of an optimal continuous-time-measurement algorithm is not guaranteed to be optimal and generally is *not*.

³For finite dimensional systems, the *state* is the set of numbers (that may change as time progresses) used to describe the system; the state and the inputs to the system determine the behavior of the system [129]. More generally, the state is an element in the smallest dimensional vector space that fully describes the system; and with knowledge of the inputs (which includes the noises and uncertainties), determines the behavior of the system.

the system model, measurement inaccuracies, and other disturbances) and thus these excised dimensions are of little value to the engineer. So, by essentially-equivalent we mean that the most essential subspace of the infinite-dimensional space is retained in the model. Additionally, unlike \mathcal{S}_1 , there is no finite element approximation to spatial differentiation, but instead a projection using a finite number of basis functions (versus the infinite number needed to describe the original function completely).

By creating a finite-dimensional model, we are able to take full advantage of the existing body of knowledge concerning digital filtering and simulation techniques and software. Finally, we use the discrete Kalman filter (\mathcal{F}_{opt}) to process the available data optimally to determine a solution recursively. The solution space might also be called the result space since not all methods are guaranteed to produce an actual solution — our process is.

1.2 *Historical Overview*

In 1960, the American Society of Mechanical Engineers (ASME) published Kalman's [95] extraordinary extension of the Wiener filter [208]; shortly thereafter, Kalman and Bucy added a second paper tackling the more mathematically sophisticated continuous-time problem, often called the Kalman-Bucy filter [96]. Kalman's filter employed a new approach that embraced signals of finite time duration, time-varying system models, and nonstationary noise processes [129] in the language of the new state space (time-domain) formulation [217, 97, 164] versus frequency-domain methods required by most Wiener filter techniques [208, 207, 129]. Kalman's work has been republished in many collections, such as [183, 12]. For a control theory point of view, see any of these excellent texts: [91, 141, 3, 129, 130, 131] or these signal processing texts: [170, 100, 77], or these newer texts that concentrate on utilizing the Kalman and related filters for tracking: [190, 193, 14]. Additionally, Kalman filtering can be viewed much as Kalman did in [95] within the mathematical frame-

work of linear algebra and functional analysis [122, 33], while Kalman-Bucy filtering [96, 32] is an especially important application of stochastic calculus [104, 159, 66].

The seeds of the multiple model technique were planted by researchers seeking to extend Kalman’s work to the case of uncertain Gaussian noise process strength using a bank of parallel Kalman filters. Magill [125] was the first to publish such an extension using a structured multiple model approach. Specifically, he investigated optimal estimation of stochastic processes which can be well modeled by a sampled Gauss-Markov process with some initially unknown yet deterministic parameters⁴; in Magill’s research, the uncertain parameters affected the statistics of the zero-mean white Gaussian dynamics driving noise. State estimation in this case is an adaptive estimation process since the estimator must adapt itself to an unknown noise environment [59]. The system chooses the “correct” parameter value from a discrete *a priori* set based on the hypothesis conditional probability calculations. Even though a parameter may vary over a continuous range of values, the MMAE method fundamentally assumes that the true parameter can be found in a discrete set. Restricting ourselves to a discrete set actually improves the performance of the MMAE since top performance occurs when the distinguishability between the elemental filters is high. When the filters “look” the same, the MMAE cannot readily “decide” which model coded in a particular elemental filter best matches the true scenario as observed in the measurements; consequently the probability flow is hampered — this topic is discussed in considerable depth in the next chapter in Sections 2.3 and 2.7.3.

⁴What is a parameter? A parameter is usually constant (or essentially constant over the time period of interest) in the form of a scalar, vector, or matrix that relates two (or more) quantities (states or signals) that vary with time (or in space). A simple deterministic parameter is the assumed-constant mass in Newton’s equation relating the force on a body undergoing acceleration: $F(t) = m a(t)$, where $F(t)$ is the time-varying force, $a(t)$ is the time-varying acceleration, and the mass m is the parameter relating the two time varying “signals.” If the parameter is not strictly constant, but varies slowly in comparison to the other quantities, we could represent the parameter by a set of values. For example, let the set of masses be expressed as a piece-wise constant function of time.

We continue our introduction to this research with a simple physically meaningful example that motivates the use of a Kalman filter and the MMAE methodology. The multiple model method effectively partitions a complex problem into adaptive and nonadaptive parts. The adaptive part is the MMAE framework that blends the estimates of several filters together to produce excellent results — usually superior to that of a single filter. The nonadaptive portion is comprised of the elemental filters themselves. While these filters could be made adaptive, this will not be pursued in this research.

1.3 Example: Navigation

To help the reader envision how an MMAE scheme improves state estimation when there are uncertainties in a subset of the parameters describing the system dynamics/measurement model or the statistics characterizing the dynamics driving noise or measurement-corruption noise, a simple example that closely resembles and then extends the “lost at sea” example by Maybeck [129] is developed in Subsections 1.3.1 and 1.3.2 respectively.

1.3.1 Kalman Filter with Known Noise Strengths. Suppose that you and a friend are sailing north from Port A to Port B. We can use a chronograph to compute our east-west position (longitude) and a sextant to determine our north-south (latitude) position. For simplicity, we assume that we have a perfect chronograph and associated charts and can thus exactly calculate our longitude. Thus you only need to find the scalar latitude position to navigate from Port A to Port B.

The one-dimensional noise that corrupts these scalar measurements obtained with a sextant by sighting the stars is an additive noise process. While you have never used your star sighting skills to navigate the sea, your friend is an accomplished expert. You take the initial star sighting to compute your latitude; let’s call it measurement $\mathbf{z}(t_1) = z_1$ and use this to establish your position at time t_1 . The

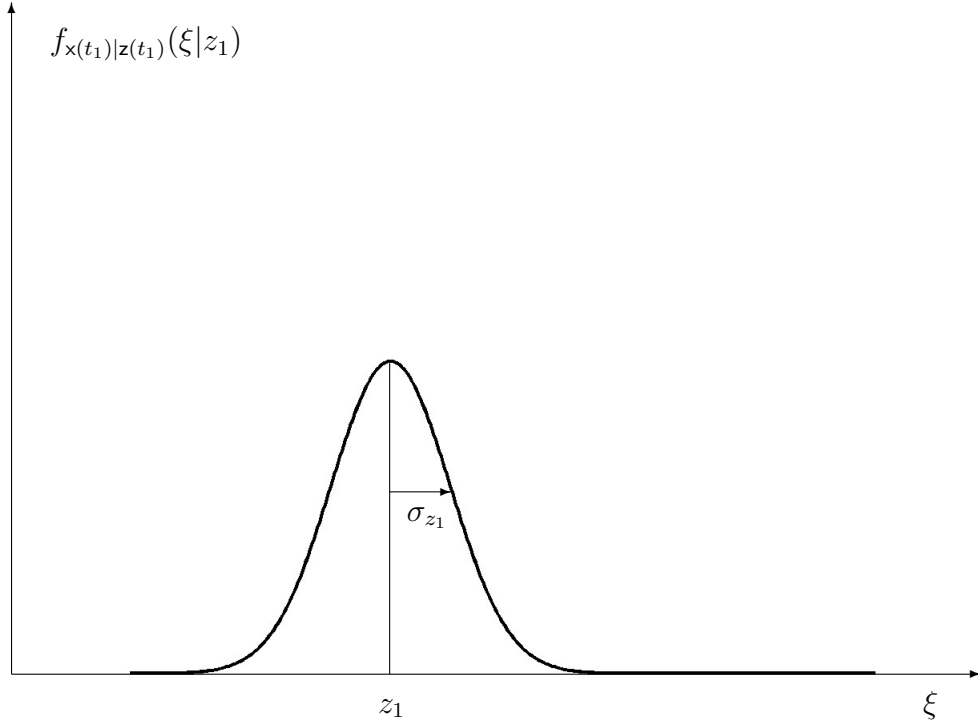


Figure 1.2 First Sextant Sighting.

precision of your measurements is characterized by the standard deviation, given by σ_{z_1} , or equivalently, the expected variance, $\sigma_{z_1}^2$, of the estimate. With this information, we can establish the conditional probability density function (PDF) of $\mathbf{x}(t_1)$, for the estimated position at time t_1 , conditioned on the observed measurement z_1 . Thus, we have two statistical moments with which to describe the conditional PDF $f_{\mathbf{x}(t_1)|\mathbf{z}(t_1)}(\xi|z_1)$. Furthermore, if we assume that this conditional PDF is Gaussian, then we have completely characterized the PDF, since only the first two moments are required to characterize the Gaussian PDF fully, as shown in Figure 1.2.

Based on the information so far, our best position (or state) estimate is

$$\hat{\mathbf{x}}(t_1) = z_1 \tag{1.3}$$

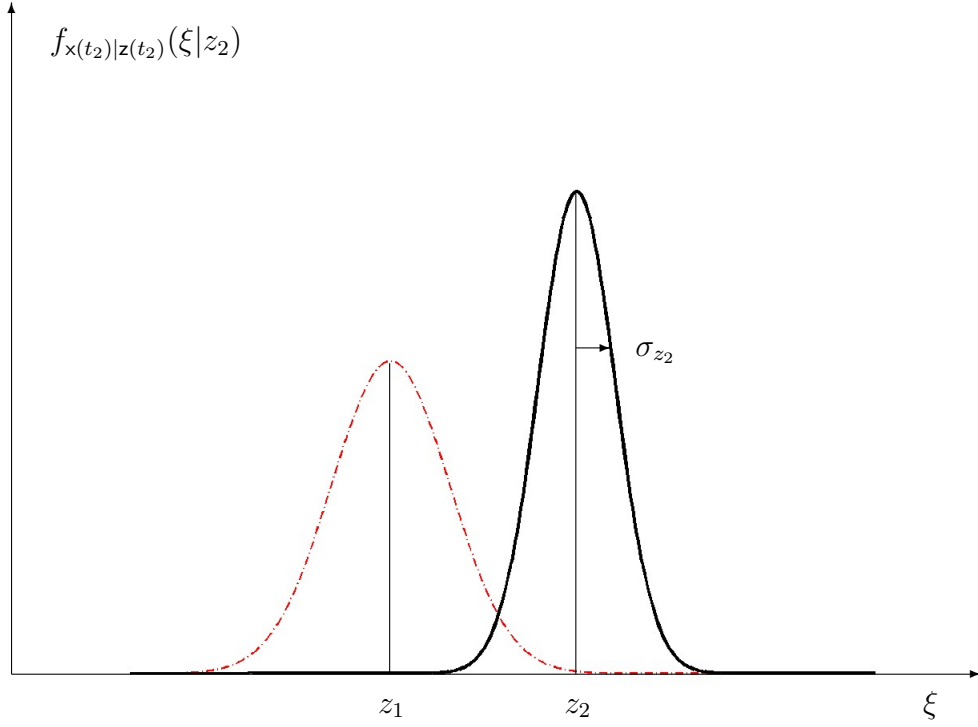


Figure 1.3 Second Sextant Sighting. The solid black line is for the second sighting and the dash-dotted gray line is for the initial sighting. The areas under the two PDFs are equivalent. The peak of the second PDF is higher since its width (characterized by the standard deviation σ_{z_2} of the PDF) is smaller than the width of the first PDF.

and the variance of the estimate error is

$$\sigma_x^2(t_1) = \sigma_{z_1}^2 \quad (1.4)$$

A few moments later, your friend takes a sighting at time $t_2 \cong t_1$, and we assume that the true position has not changed at all — this is equivalent to taking the two measurements at the same time with two identical sextants. Label his measurement z_2 with error variance $\sigma_{z_2}^2$. Note that the difference in the models is confined to the assessed skill of the sextant operator. Figure 1.3 shows conditional PDFs for both measurements.

Now we have two measurements with which to estimate our position. How do we combine them to yield the best estimate? If we didn't have any knowledge about the precision of the measurements, we would simply average the measurements to estimate the position at time t_2

$$\hat{x}(t_2) = \frac{1}{2}z_1 + \frac{1}{2}z_2 \quad (1.5)$$

Observe that the coefficients sum to one. Since we know the precision of each measurement — in terms of the expected variance — let's use them. A clever person might propose a weighted average of the measurements z_1 and z_2 in terms of the expected variances to yield a position estimate at time t_2 of [17]

$$\hat{x}(t_2) = \frac{\sigma_{z_2}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_1 + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_2 \quad (1.6)$$

where the coefficients yet again sum to one and the expected variance of the position estimate conditioned on the accuracy of the measurements is

$$\sigma_x^2(t_2) = \frac{1}{(1/\sigma_{z_1}^2) + (1/\sigma_{z_2}^2)} \quad (1.7)$$

In fact, it can be shown that Equations (1.6) and (1.7) correspond to the conditional mean and error covariance of a Gaussian PDF conditioned on the two measurements [129]. Given this conditional Gaussian PDF, this estimate is *optimal* under many criteria, such as the minimum mean-squared error — see Chapter II in general and Section 2.3.2 in particular. The mean is simply a weighted average of the two measurements and the variance is reduced by adding a second measurement — regardless of the precision of or the variance associated with the second measurement, assuming that the variance is not infinite. Curiously, the equation for the new variance is of the same structure as that of adding two resistors in parallel. Next we substitute

Equation (1.3) into Equation (1.6) and rewrite it in a “predictor-corrector” form

$$\hat{x}(t_2) = \hat{x}(t_1) + K(t_2)[z_2 - \hat{x}(t_1)] \quad (1.8)$$

where $K(t_2)$, the so-called Kalman gain, which is “chosen” to minimize the mean-squared error between the position estimate and the true position, is given by

$$K(t_2) = \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} \quad (1.9)$$

Similarly, the variance in Equation (1.7) can be rewritten using Equation (1.9) with Equation (1.4) as

$$\sigma_x^2(t_2) = \sigma_x^2(t_1) - K(t_2) \sigma_x^2(t_1) \quad (1.10)$$

Thus, the Gaussian conditional PDF $f_{x(t_2)|z(t_1),z(t_2)}(\xi|z_1, z_2)$ at time t_2 is completely specified by mean $\mu = \hat{x}(t_2)$ and variance $\sigma^2 = \sigma_x^2(t_2)$ shown in Equations (1.8) and (1.10) respectively, with the Kalman filter gain defined in Equation (1.9) and displayed in Figure 1.4.

Now that we have shown how to update the initial estimate with a second measurement, let’s add some dynamics to the problem in order to estimate a future position *before* we take a third measurement. Basic kinematics enables us to model your change in position as

$$dx/dt = u + w \quad (1.11)$$

where dx/dt represents the rate of change in position or velocity, u is some nominal velocity, and the uncertainty in the velocity due to unmodeled effects and other disturbances is described with a zero-mean Gaussian random variable w , often simply referred to as “noise.” This noise is uncorrelated in time (i.e., it is a white process) and has strength Q ; the strength Q of the noise corresponds to the amplitude of the power spectral density (PSD) curve over all frequencies, which for this white noise process is constant for all time. With this simple model, we can predict the position

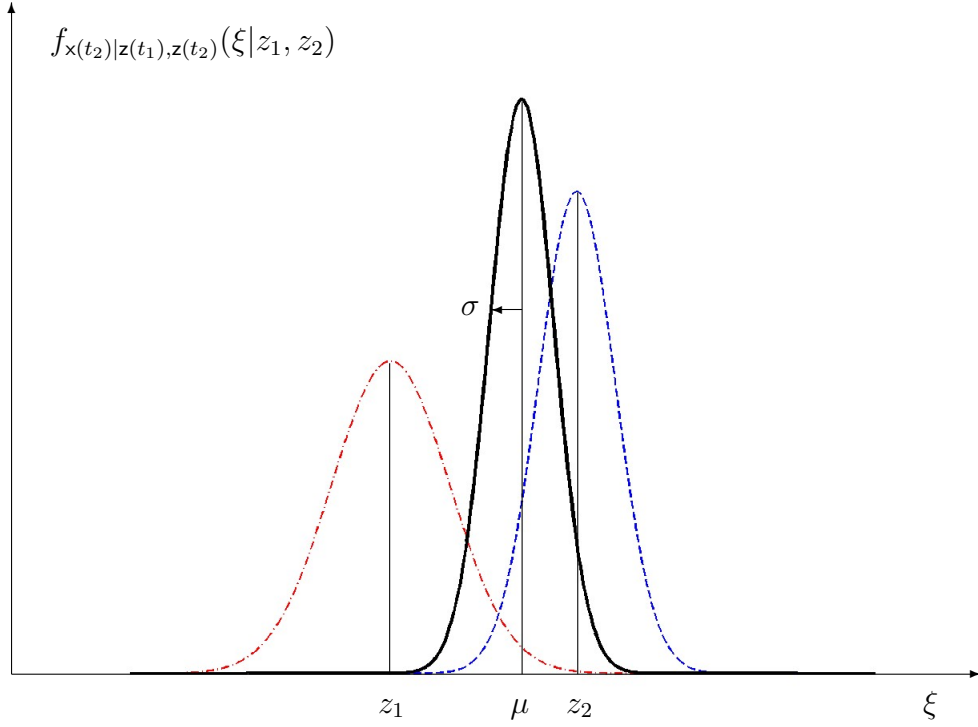


Figure 1.4 First and Second Sextant Sightings Combined. Solid black line is the combination of the two sightings, the dashed gray line is for the second sighting, and the dash-dotted gray line is for the first sighting.

at time t_3 as

$$\hat{x}(t_3^-) = \hat{x}(t_2) + u[t_3 - t_2] \quad (1.12)$$

where the time t_3^- is a notational convenience representing our *prediction* at time t_3 before a new measurement is added and the variance of the estimate grows to

$$\sigma_x^2(t_3^-) = \sigma_x^2(t_2) + Q[t_3 - t_2] \quad (1.13)$$

With this construction, we have in effect *propagated* the Gaussian conditional PDF from time t_2 until time t_3 , which we represent as t_3^- in order to keep track of our propagation and update stages. For the linear dynamics model in Equation (1.11) driven by known inputs u and white Gaussian noise w , the PDF will continue to be Gaussian [129].

Next, we add a third measurement taken by your friend and update the position estimate. This update will complete the second step of this two-stage process that consists of *propagating* the optimal estimate forward in time using the dynamics model and then *updating* the estimate with a new measurement. Note that the variance of our estimate grows during the propagation stage for this problem because of the uncertainty in our initial position estimate $\hat{x}(t_1)$ and the uncertainty in how our position changes over time (due to the \mathbf{w}). On the other hand, the variance is reduced during the update stage because we have added new information to refine the position estimate. Using Equations (1.8) and (1.10) with the Kalman filter gain defined in Equation (1.9) as our template, we write the optimal position prediction at time t_3 as

$$\hat{x}(t_3) = \hat{x}(t_3^-) + K(t_3)[z_3 - \hat{x}(t_3^-)] \quad (1.14)$$

with variance

$$\sigma_x^2(t_3) = \sigma_x^2(t_3^-) - K(t_3) \sigma_x^2(t_3^-) \quad (1.15)$$

where the Kalman gain is

$$K(t_3) = \frac{\sigma_x^2(t_3^-)}{\sigma_x^2(t_3^-) + \sigma_{z_3}^2} \quad (1.16)$$

Hence, the optimal estimate of the position x at time t_3 is the optimal predicted position just before the latest measurement, $\hat{x}(t_3^-)$, plus a correction term based on the weighted residual between the new measurement and the measurement prediction $K(t_3)[z_3 - \hat{x}(t_3^-)]$, where the Kalman gain $K(t_3)$ provides the weighting and $[z_3 - \hat{x}(t_3^-)]$ is called the residual. The weighting represents our relative confidence in the measurements and predicted position estimates. We can continue this process of propagating and updating indefinitely until our objective (reaching Port B) is achieved. If our sighting variances $\sigma_{z_i}^2$, for each measurement are accurate, then we can expect good results from this single Kalman filter. However, if there is some uncertainty in the declared statistics (and/or models), then the performance generally degrades from what we would otherwise expect.

1.3.2 Multiple Model Adaptive Estimation. Now that we have shown how a Kalman filter can assist our navigation solution by improving our latitude estimate, we shall consider the case for which there is an unpredictable uncertainty with the quality of our sextant measurements. Specifically, we will consider a system characterized by a known dynamics noise strength and two possible measurement noise covariances in order to motivate the utility of a multiple model estimator. For our example, these two “choices” represent the measurement-corruption variances for a properly calibrated sextant and another for an uncalibrated sextant. This problem is similar to the problem that Magill [125] studied, except that he did not limit himself to just two possibilities and he was concerned with an uncertain dynamics driving noise.

Once again, you and a friend are sailing from Port A to Port B on a clear night. You take the initial sighting using the calibrated sextant and determine your latitude $z(t_1) = z_1$ at time t_1 with variance $\sigma_{z_1}^2$. As you hand the sextant to your friend to take the remainder of the measurements, it slips from your grip and falls to the deck. You are uncertain if this fall has spoiled the calibration of your instrument; however, you have a measurement variance recorded in your log book for uncalibrated sextants. So, you are faced with more uncertainty in your measurements...or are you? The question remains whether the measurement variance for the subsequent sightings is equal to either R_{cal} or R_{uncal} , where $R_{\text{cal}} < R_{\text{uncal}}$. But don't despair, the MMAE can tell you whether the sextant is most likely calibrated or not after just a few more measurements and it can give you the best latitude estimate possible. The MMAE technique is well suited to give the proper “advice.”

In this simple example, we shall operate two Kalman filters F_1 and F_2 in parallel (see Figure 1.5) — one for a calibrated sensor and one for an uncalibrated sextant — each processing the same measurements, $z(t_i) = z_i$, and each propagating the estimate using the same algorithm developed in Section 1.3.1. In block D we *determine* the probability, p_j , that elemental filter F_j is the best modeled filter. The

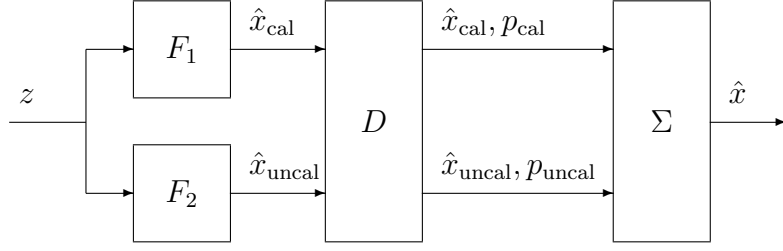


Figure 1.5 Multiple Model Adaptive Estimation: Two Kalman Filters in Parallel
 Σ block processes the estimates from the elemental filters and creates an overall position estimate.

Let's begin by building two filters and label the first filter “cal” for a properly calibrated sensor and the second, “uncal” for an uncalibrated sensor. Recall that you performed the first measurement using the calibrated sextant, call it $z(t_1) = z_1$ at time t_1 . The initial position estimate is the same for both the “cal” and “uncal” filters

$$z_1 = \hat{x}_{\text{cal}}(t_1) = \hat{x}_{\text{uncal}}(t_1) \quad (1.17)$$

with the same initial *measurement-corruption noise covariance*

$$\sigma_{z_1}^2 = R_{\text{cal}} \quad (1.18)$$

since it was taken using the calibrated sextant before it was dropped. Using Equations (1.12) and (1.13) the position estimate at time t_i is

$$\hat{x}_{\text{cal}}(t_i^-) = \hat{x}_{\text{cal}}(t_{i-1}) + u[t_i - t_{i-1}] \quad (1.19)$$

$$\hat{x}_{\text{uncal}}(t_i^-) = \hat{x}_{\text{uncal}}(t_{i-1}) + u[t_i - t_{i-1}] \quad (1.20)$$

with variances for the calibrated and uncalibrated models

$$\sigma_{\text{cal}}^2(t_i^-) = \sigma_{\text{cal}}^2(t_{i-1}) + Q[t_i - t_{i-1}] \quad (1.21)$$

$$\sigma_{\text{uncal}}^2(t_i^-) = \sigma_{\text{uncal}}^2(t_{i-1}) + Q[t_i - t_{i-1}] \quad (1.22)$$

and the *dynamics-process noise strength*, Q , is assumed constant for all time t_i .

Next, the two filters are updated at time t_i using the latest measurement $z(t_i) = z_i$ following the form of Equations (1.14) through (1.16) we obtain

$$\hat{x}_{\text{cal}}(t_i) = \hat{x}_{\text{cal}}(t_i^-) + K_{\text{cal}}(t_i)[z_i - \hat{x}_{\text{cal}}(t_i^-)] \quad (1.23)$$

$$\hat{x}_{\text{uncal}}(t_i) = \hat{x}_{\text{uncal}}(t_i^-) + K_{\text{uncal}}(t_i)[z_i - \hat{x}_{\text{uncal}}(t_i^-)] \quad (1.24)$$

with variances

$$\sigma_{\text{cal}}^2(t_i) = \sigma_{\text{cal}}^2(t_i^-) - K_{\text{cal}}(t_i) \sigma_{\text{cal}}^2(t_i^-) \quad (1.25)$$

$$\sigma_{\text{uncal}}^2(t_i) = \sigma_{\text{uncal}}^2(t_i^-) - K_{\text{uncal}}(t_i) \sigma_{\text{uncal}}^2(t_i^-) \quad (1.26)$$

and Kalman gains of

$$K_{\text{cal}}(t_i) = \sigma_{\text{cal}}^2(t_i^-) [\sigma_{\text{cal}}^2(t_i^-) + R_{\text{cal}}]^{-1} \quad (1.27)$$

$$K_{\text{uncal}}(t_i) = \sigma_{\text{uncal}}^2(t_i^-) [\sigma_{\text{uncal}}^2(t_i^-) + R_{\text{uncal}}]^{-1} \quad (1.28)$$

where the measurement precision is quantified by the assumed constant variances of the corruption noises R_{cal} and R_{uncal} for all time t_i . The difference between the predicted measurement, $\hat{x}_{\text{cal}}(t_i^-)$ or $\hat{x}_{\text{uncal}}(t_i^-)$, and the actual observation, z_i for all time t_i , is termed the residual in each of the two elemental filters. The residual, $[z_i - \hat{x}_{\text{cal}}(t_i^-)]$ or $[z_i - \hat{x}_{\text{uncal}}(t_i^-)]$, is our connection to the “real” world in Equations (1.23) and (1.24), respectively.

Now that we have our two *elemental* filters set up, how do we decide which one is the most accurate so that we can form the best estimate of our position? In this development, we have implicitly assumed that the sextant is either calibrated or uncalibrated and that one of the filters will give us the best results. The probability, $\text{pr}\{\cdot\}$, that a hypothesis is true given the observations is termed the hypothesis conditional probability. The hypothesis conditional probability for the calibrated sextant, $p_{\text{cal}}(t_i)$, and uncalibrated sextant, $p_{\text{uncal}}(t_i)$, are defined as

$$p_{\text{cal}}(t_i) \triangleq \text{pr}\{\mathbf{R} = R_{\text{cal}} | \mathbf{z}(t_1) = z_1, \mathbf{z}(t_2) = z_2, \dots, \mathbf{z}(t_i) = z_i\} \quad (1.29)$$

$$p_{\text{uncal}}(t_i) \triangleq \text{pr}\{\mathbf{R} = R_{\text{uncal}} | \mathbf{z}(t_1) = z_1, \mathbf{z}(t_2) = z_2, \dots, \mathbf{z}(t_i) = z_i\} \quad (1.30)$$

such that $p_{\text{cal}}(t_i), p_{\text{uncal}}(t_i) \geq 0$, and they sum to one: $p_{\text{cal}}(t_i) + p_{\text{uncal}}(t_i) = 1$ for every time t_i . We introduce the following shorthand notations for the PDFs for the measurement at time t_i conditioned on the assumed parameter value for \mathbf{R} and the sequence of observations from time t_1 through time t_{i-1} — this is also known simply as the conditional PDF for the incoming measurement and is given here as [130]

$$f_{\text{cal}}(t_i) \equiv f_{\mathbf{z}(t_i) | \mathbf{R}, \mathbf{z}(t_1), \dots, \mathbf{z}(t_{i-1})}(z_i | R_{\text{cal}}, z_1, \dots, z_{i-1}) \quad (1.31)$$

$$f_{\text{uncal}}(t_i) \equiv f_{\mathbf{z}(t_i) | \mathbf{R}, \mathbf{z}(t_1), \dots, \mathbf{z}(t_{i-1})}(z_i | R_{\text{uncal}}, z_1, \dots, z_{i-1}) \quad (1.32)$$

These PDFs are a function of the residual, $[z_i - \hat{x}_{\text{cal}}(t_i^-)]$ or $[z_i - \hat{x}_{\text{uncal}}(t_i^-)]$, as seen in Equations (1.23) and (1.24), and the filter-computed residual variance, $[\sigma_{\text{cal}}^2(t_i^-) + R_{\text{cal}}]$ or $[\sigma_{\text{uncal}}^2(t_i^-) + R_{\text{uncal}}]$, which appear in Equations (1.27) and (1.28). Thus, these PDFs contain the information we use to calculate the hypothesis conditional probabilities. These probabilities indicate how well the hypothesized models, $\mathbf{R} = R_{\text{cal}}$ for the calibrated sextant filter model and $\mathbf{R} = R_{\text{uncal}}$ for the uncalibrated sextant filter model, match the real world. We judge the quality of the match using the sequence of residuals created from the predicted measurements $\hat{x}_{\text{cal}}(t_i^-)$ and $\hat{x}_{\text{uncal}}(t_i^-)$ and the observed measurements z_i for all time t_i . The filter model that

best matches the reality will produce the “whitest” sequence of residuals⁵, a mean that is closest to zero, and a residual sequence variance that is most in consonance with the filter-computed variance, $[\sigma_{\text{cal}}^2(t_i^-) + R_{\text{uncal}}]$ or $[\sigma_{\text{uncal}}^2(t_i^-) + R_{\text{uncal}}]$. This will, in turn, increase the value of the measurement conditional PDF evaluated for a measurement at time t_i , which in turn leads to a higher hypothesis conditional probability that this particular model matches the real world the best. Additionally, the measurement uncertainties are visible in the Kalman gain Equations (1.27) and (1.28). Furthermore, it can be shown that these measurement conditional PDFs are Gaussian with a mean equal to the predicted state estimate, $\hat{x}_{\text{cal}}(t_i^-)$ or $\hat{x}_{\text{uncal}}(t_i^-)$, and a known variance $[\sigma_{\text{cal}}^2(t_i^-) + R_{\text{cal}}]$ or $[\sigma_{\text{uncal}}^2(t_i^-) + R_{\text{uncal}}]$.

The measurement conditional PDFs can be evaluated for a particular assumed R , and thus $f_{\text{cal}}(t_i)$ and $f_{\text{uncal}}(t_i)$ are evaluated as numbers once $\mathbf{z}(t_i) = z_i$ becomes available. Then, the probability that the sextant is calibrated or uncalibrated at time $t_i > t_1$ is⁶

$$p_{\text{cal}}(t_i) = \frac{f_{\text{cal}}(t_i) p_{\text{cal}}(t_{i-1})}{f_{\text{cal}}(t_i) p_{\text{cal}}(t_{i-1}) + f_{\text{uncal}}(t_i) p_{\text{uncal}}(t_{i-1})} \quad (1.33)$$

$$p_{\text{uncal}}(t_i) = \frac{f_{\text{uncal}}(t_i) p_{\text{uncal}}(t_{i-1})}{f_{\text{cal}}(t_i) p_{\text{cal}}(t_{i-1}) + f_{\text{uncal}}(t_i) p_{\text{uncal}}(t_{i-1})} \quad (1.34)$$

These two probabilities are nonnegative and the denominator in Equations (1.33) and (1.34) serves to scale the probabilities so that they sum to one. For the first measurement, we assumed that the sextant was calibrated; thus we had $p_{\text{cal}}(t_1) = 1$ and $p_{\text{uncal}}(t_1) = 0$. Using Equations (1.33) and (1.34), we can calculate the hypothesis

⁵A sequence of white residuals is one of the criteria that indicates that the assumed filter model is in consonance with the real world or with the truth model in the case of a simulation [129] as discussed in Section 2.3.3.3.

⁶Recall that we assumed that the sextant was initially calibrated; thus $p_{\text{cal}}(t_1) = 1$ and $p_{\text{uncal}}(t_1) = 0$.

conditional probabilities at time t_2 as

$$p_{\text{cal}}(t_2) = \frac{f_{\text{cal}}(t_i) \cdot 1}{f_{\text{cal}}(t_i) \cdot 1 + f_{\text{uncal}}(t_i) \cdot 0} = \frac{f_{\text{cal}}(t_i)}{f_{\text{cal}}(t_i)} = 1 \quad (1.35)$$

$$p_{\text{uncal}}(t_2) = \frac{f_{\text{uncal}}(t_i) \cdot 0}{f_{\text{cal}}(t_i) \cdot 1 + f_{\text{uncal}}(t_i) \cdot 0} = \frac{0}{f_{\text{cal}}(t_i)} = 0 \quad (1.36)$$

Thus, $p_{\text{cal}}(t_i) = 1$ and $p_{\text{uncal}}(t_i) = 0$ for all time t_i . This presents a problem with a known remedy. An elemental filter is virtually removed from the filter bank if the hypothesis conditional probability goes to zero. By inspection of Equations (1.33) and (1.34), we see that the probability at the next instant of time can no longer change after it becomes zero. A popular method used to counteract the “lock-out” problem attributed to this particular method used to compute the probabilities using Equations (1.33) and (1.34) is to introduce some additional logic into the algorithm. One technique that has been used with success imposes a lower bound⁷ on the value that the hypothesis conditional probabilities may assume, thus prohibiting it from being driven to zero. We shall choose one tenth as our lower bound for this example. Thus we need to recompute the calibrated and uncalibrated hypothesis conditional probabilities given in Equations (1.35) and (1.36) using $p_{\text{cal}}(t_1) = 0.9$ and $p_{\text{uncal}}(t_1) = 0.1$. Hence,

$$p_{\text{cal}}(t_2) = \frac{f_{\text{cal}}(t_i) \cdot 0.9}{f_{\text{cal}}(t_i) \cdot 0.9 + f_{\text{uncal}}(t_i) \cdot 0.1} \quad (1.37)$$

$$p_{\text{uncal}}(t_2) = \frac{f_{\text{uncal}}(t_i) \cdot 0.1}{f_{\text{cal}}(t_i) \cdot 0.9 + f_{\text{uncal}}(t_i) \cdot 0.1} \quad (1.38)$$

If $p_{\text{cal}}(t_2) \geq 0.1$ and $p_{\text{uncal}}(t_2) \geq 0.1$, then we have acceptable probabilities and we are done with this time step. However, if either inequality fails to be true, then we round the too-low probability up to 0.1 and round the other probability down to 0.9. Note that this lower bounding process becomes more involved when we have more than two elemental filters: once the lower bounds are imposed, all probabilities must

⁷See Section 2.4.6 for more information on lower bounding the hypothesis conditional probability.

be rescaled to ensure that their sum is one. We then proceed to time t_3 in the same fashion.

We compute the hypothesis conditional probabilities to determine whether the sextant is either calibrated or uncalibrated. If the drop did not affect the calibration of the sextant, then, it is likely that $p_{\text{cal}}(t_i) > p_{\text{uncal}}(t_i)$ and the best position estimate will use the “cal” filter since it has the highest hypothesis conditional probability. But since our two models were chosen rather arbitrarily from a continuous set of possibilities, a blended estimate based on the position estimates from both elemental filters might be best approach, with blending weights given by the hypothesis conditional probabilities. Note that the parameter estimate will fall at the end points or in between the values used in the models; i.e., we don’t extrapolate outside of the filter bank. Thus, choosing a small R_{cal} value for a tightly calibrated sextant and a high R_{uncal} value for a *very uncalibrated* sextant might produce the best results.

The Bayesian state estimate is readily computed via the MMAE. The Bayesian estimate represents the minimum mean-squared error blending of information provided by the MMAE in the form of position estimates and hypothesis conditional probabilities for each elemental filter. The hypothesis conditional probabilities are used to weight each position estimate as shown

$$\hat{x}_{\text{MMAE}}(t_i) = \hat{x}_{\text{cal}}(t_i) p_{\text{cal}}(t_i) + \hat{x}_{\text{uncal}}(t_i) p_{\text{uncal}}(t_i) \quad (1.39)$$

We could also form an estimate of the measurement variance based on these two filters, if for instance the sextant was slightly damaged and thus partially uncalibrated, this might reveal itself in the measurement residuals and thus

$$\hat{R}_{\text{MMAE}}(t_i) = R_{\text{cal}} p_{\text{cal}}(t_i) + R_{\text{uncal}} p_{\text{uncal}}(t_i) \quad (1.40)$$

would give us a better estimate than either filter by itself.

In this section we have shown how two Kalman filters can be used in the MMAE framework to supply positional estimates needed to navigate north to the Port B latitude better than one filter when the precision of the sensor measurements is in doubt. The derivations to some of these equations will be undertaken in the next chapter, while most are simply stated with reference to where the reader may find the derivation. An overview of the advanced topics used in this research is next.

1.4 *Advanced Topics Overview*

The linear systems theory state space approach [217, 164, 155] for solving systems of linear differential equations was just taking root when Kalman [95] reported his new approach for linear filtering and prediction over forty years ago. These foundations have been further refined and extolled by many researchers, in engineering and applied mathematics, studying the complementary fields of mathematical optimization, information theory, signal processing, estimation, system identification, and control, such as Kalman, Bucy, Falb, Arbib, Sorenson, and others [97, 183, 182, 6]. In this research, we will be mainly interested in the extensions of linear systems filtering theory to infinite dimensions [38, 18, 19, 39]⁸, since all real physical systems are truly distributed [163, 184, 185]. The bulk of the work on infinite-dimensional systems theory has been reported for specific problem types, such as the solution to a parabolic PDE, see for example [161, 18, 19], while in this research, we make no assumptions on the particular type of infinite-dimensional system during our derivation of the ISKF. However, discussions of topics such as observability are largely problem-specific and usually apply to subclasses of problems such as various types of PDEs [38, 163, 184, 185, 18, 19]. We will not report any results with regard to the very important topic of observability or its dual, controllability. Other researchers have extended the study of infinite-dimensional linear

⁸Other than the seminal paper by Kalman, most of the references cited in this section are simply good references that have been used by the author during this research and may or may not be (although they sometimes are) the first or definitive work on the subject.

systems theory to functional differential equations, as reported in [103]. Additionally, the semigroup theory has been steadily developed to characterize the solutions to a multitude of initial value or abstract Cauchy problems (ACP) that feature the evolution equation [83, 216, 160, 48]. During the past thirty years, mathematical control theorists have studied and reported on the evolution equation formulation [38, 160, 39]. Probability theory, stochastic processes, and stochastic calculus, as developed in [42, 40, 66, 45], are needed to characterize fully the additive noise processes perturbing our evolution equations representing the dynamics model as well as the measurement model equations. Finally, our development of ISKF is framed and executed using the tools of functional analysis [83, 154, 36, 37, 24].

While Chapter II gives a thorough background for the MMAE, it provides only brief descriptions of some of the advanced material mentioned above. All of the advanced topics given above are developed to the level necessary for the derivation of the ISKF and the algorithm for creating the equivalent infinite-dimensional discrete-time model in Chapter III. Furthermore, we employ a Galerkin-like method⁹ to create an essentially-equivalent discrete-time model for our stochastic PDE problem in Chapter IV. But first, we conclude this chapter with some notes to the reader, a summary, and a fuller outline of the rest of the dissertation.

1.5 Notes to the Reader

1. References: When multiple references are cited in the text, the ordering is generally in chronological order from the first source to the most recent. The primary exceptions to this rule occurs when a source is difficult to obtain and a more recent source adequately explains the topic (or sometimes more clearly than the original source) and gives proper credit to the original source; such a source will be listed first. This will usually occur without further clarification.

⁹See [63, 89, 62, 30, 61] for an accounting of the Galerkin method that has been used to find approximate solutions to PDEs.

On the other hand, the bibliography cites the references in strictly alphabetical order for each author and chronologically for each author with multiple contributions.

2. Notations: Due to the confluence of several engineering and mathematical notational conventions, we have at times created a new or slightly altered symbol for a common quantity or concept so that a single symbol rarely has multiple meanings that must be ascertained from the context. Oftentimes the font type is used to differentiate two similar symbols. For the most part, the notational convention used by Maybeck [129, 130] is used; e.g., \mathbf{x} is a random state variable as indicted by the sans serif font used, while a realization of \mathbf{x} is denoted by $\mathbf{x}(\omega) = x$. A second example calls on the Hilbert space, \mathbb{H} , the measurement distribution operator, H , and the measurement distributor matrix, \mathbf{H} ; all are typeset using upper case lettering, but with different font types. Another good example is the set of real numbers, \mathbb{R} , the measurement residual vector, \mathbf{r} , the measurement noise covariance matrix, \mathbf{R} , the range of a mapping, \mathcal{R} , and the random measurement noise covariance \mathbf{R} we just saw in the navigation example. In addition to the general rules that follow, see the complete list of symbols beginning on page xvii.

▷ *Scalars* are denoted by both upper and lower case letters in italic type for Arabic letters and lower case only for Greek letters. For example, j , N , and κ are all scalars.

▷ *Vectors* are denoted by lower case letters in boldface type, such as \mathbf{x} or $\boldsymbol{\alpha}$.

▷ *Matrices* are denoted by upper case letters in boldface type like \mathbf{R} and $\boldsymbol{\Phi}$. Additionally, an $n \times m$ matrix, for $m = 1$ or $n = 1$, while technically a vector or the transpose of a vector, will sometimes retain the boldface *upper* case typesetting when there is more to gain by using the familiar typeset.

▷ *Functions* are generally denoted in the same fashion as are scalars; however, when a function is treated as an element in an infinite-dimensional vector space,

it is often written as a vector in boldface type. For example, f and \mathbf{x} are both functions; f is a scalar function, while \mathbf{x} is generally an n -vector of functions; however, n could be one.

▷ *Transformations and operators* are denoted by upper case letters in italic, as in F for transformations (and operators) that are associated with Kalman filtering, and calligraphy type script for other ‘standard’ operators such as the conditional expectation operator \mathcal{E} . The only reason for this difference is simply to maintain the aesthetics of the notation developed for finite-dimensional filtering.

▷ *Sets* are denoted by upper case double-lined blackboard type: \mathbb{A} and \mathbb{X} . Special sets such as the σ field, \mathcal{F} , are usually denoted using the calligraphy font.

▷ *Set operators* are often denoted by a second calligraphy type. Two examples are \mathfrak{B} and \mathfrak{F} .

▷ *Random vectors* and *vector stochastic processes* are set in boldface sans serif type, e.g., \mathbf{x} .

▷ *Random variables* and *scalar stochastic processes* are set in sans serif type, as in x .

▷ *Realizations* of the random vector are set in boldface roman type, $\mathbf{x}(\omega_i) = \mathbf{x}$, while its scalar components (realizations of random variables) are denoted in italics as x_k , for $k = 1, 2, \dots$

▷ Similarly, *samples* of the stochastic vector process are set in boldface roman type, $\mathbf{x}(t, \omega_i) = \mathbf{x}(t)$, while its scalar components (samples of stochastic scalar processes) are denoted by $x_k(t)$, for $k = 1, 2, \dots$

▷ Finally, a few special operators are given in standard form: the integral, $\int_a^b x(t) dt$; the sum, $\sum_{i=1}^N x_i$; the product, $\prod_{i=1}^N x_i$; the intersection, $\bigcap_{i=1}^N A_i$.

1.6 Summary

In this chapter we introduced the Kalman filter and the multiple model adaptive estimation (MMAE) technique which is used to extend Kalman filtering in two manners. Using a bank of elemental filters, the MMAE method can improve state estimation by readily adapting to an unknown noise environment or by identifying uncertain system parameters. Additionally, it can be tuned to focus on performing system identification (also known as parameter estimation) in either a known or unknown noise environment. Previous researchers have concentrated on finite-dimensional problems; this research expands the class of problems to encompass problems more accurately modeled using infinite-dimensional state space descriptions. Much of the previous work on distributed-parameter systems and systems featuring time-delayed differential equation models have relied on *ad hoc* methods to solve their problems. This research puts these sorts of problems on firm theoretical ground. While one must make approximations at some point to produce an algorithm to run on a digital computer, these approximations occur later in the design process presented herein, and they can be optimized for the computational load or whatever criteria is most important for the application.

1.7 The Rest of the Dissertation

In Chapter II we give a lengthy accounting of MMAE techniques. While only a small portion of this extensive review is necessary for understanding the fixed-bank MMAE method employed in this research, it was included as a useful survey for the reader interested in extending this line of research where more robust moving-bank variants of MMAE are essential to improved estimation performance or in response to tighter restrictions on the computational loading.

In Chapter III we derive the infinite-dimensional sampled-data Kalman filter (ISKF). We also develop a method (analogous to the technique used for finite-dimensional systems as described in the previous chapter) for creating the equivalent

infinite-dimensional discrete-time model from the infinite-dimensional continuous-time model. Finally, we extend the structure of fixed-bank MMAE framework so that it may accept the ISKF, thus creating the generalized infinite-dimensional MMAE (GIMMAE).

In Chapter IV we demonstrate our new and modified techniques in an extended example using the stochastic heat equation. To demonstrate the power and advantages of the methodology developed in the previous chapter, we devote a considerable portion of this work to simulating, in Chapter V, the state estimation performance of an MMAE populated with Kalman filters based on the essentially-equivalent finite-dimensional discrete-time model we created. The MMAE demonstrates good state estimation performance while operating in an uncertain noise environment; additionally, the MMAE is shown to be capable of quickly performing system identification.

In Chapter VI we begin by reviewing the contributions of this research. Next, we draw some conclusions, and finally, we offer recommendations for future work.

II. Multiple Model Adaptive Estimation

2.1 Introduction

As we saw in Chapter I there are many problems which can be adequately modeled by stochastic differential equations, upon which a Kalman filter or set of Kalman filters could be based. When there are uncertain model parameters, a group of Kalman filters, such as a parallel bank of filters, acting in concert, generally provides a better state estimate than a single filter. In addition to providing a superior state estimate, the filter bank structure provides good estimates of the uncertain model parameters. To that end, we endeavor to describe and discuss the estimation framework known as multiple model adaptive estimation (MMAE). Additionally, many enhancements to the basic structure are presented in this chapter. Multiple model methods have been employed in the following areas: target tracking [140, 14, 137], aided inertial navigation systems [50, 144], sensor and actuator failure detection and identification [56], aircraft and space structures (guidance and control) [65, 171], drug infusion¹ [127, 215], and chemical process control [173].

Only the first four sections of this background chapter are necessary to understand the formative stages and the basic fundamental concepts of MMAE as applied in this research. Following this introductory section, we give a synopsis of the early contributions that laid the groundwork for today's MMAE research. Next, the structure and components of MMAE are presented. The final section of essential reading contains an extensive collection of practical performance enhancements to improve the performance of the MMAE. Section 2.5 provides an excellent example on how a single filter relates to a bank of filters. Lastly, Section 2.6 introduces a class of

¹Drug infusion is a diffusion process with time-varying parameters; hence it is one of the areas that may benefit from the infinite-dimensional approach developed in this research.

dynamic filter bank techniques used both to increase the breadth of the filter bank while retaining the resolution of a narrowly focused static bank of filters.

2.2 *The Beginnings of Multiple Model Adaptive Estimation and Control*

In this section, we highlight the first twenty years of MMAE-related research appearing in the literature. This research laid the groundwork for the multiple model methodology that has blossomed over the past forty years since Magill [125] proposed his method for employing a set of Kalman filters [95] in an uncertain noise environment. While we have assumed a certain familiarity with Kalman filtering, we have described and discussed in-depth the dynamics and measurement models², as well as the filtering algorithm itself. A significant cross-section of the available literature is cited as we endeavor to prepare a solid foundation for a deep coverage of multiple model methodology and associated techniques that are the main subjects of this chapter. This first subsection is organized chronologically to emphasize the growth and increasing sophistication of the research area over time.

In 1965, Magill [125] presented a novel state estimation technique for a sampled Gauss-Markov stochastic process. He employed multiple models to address the problem of unknown parameter variation from within a finite set of known values. For this inaugural work, the parameters described the statistics of the dynamics driving noise, which for a time-invariant system model are stationary. Each of the parameter values corresponded to a hypothesis of the real world with a stationary noise process and was used to construct an elemental filter. These hypotheses were tested by computing the conditional probability of each hypothesis being correct. These probabilities are conditioned on the observed measurements for each filter. Magill established the notion that a random “switch” selects the elemental stochastic process that is in force. Using the hypothesis conditional probabilities as weights, he

²In the estimation research discussed in this dissertation, the operation of the system or plant is governed by a *dynamics* model, while the system behavior is imperfectly observed using a *measurement* model.

formed a weighted sum of the state estimates from each filter; therefore the estimate is a blend of all of the models hypothesized to represent the real world system. This estimate is optimal provided that one of the assumed models matches the physical process when the unknown parameter is a constant vector. Furthermore, he asserted that a sufficient condition for this optimality is: if all of the elemental stochastic processes are ergodic, then the weighting coefficients will converge with probability one to unity for the true process and to zero for the others.

Early in 1969, Hillborn and Lainiotis [82] extended Magill's work from the case of scalar measurements to the vector measurement case and presented an optimal conditional mean estimator based on unknown (but constant) parameters. They state that, under certain necessary and sufficient conditions, in a Bayesian sense, the optimality of their state estimate is independent of the convergence to the value of the unknown parameter. Hence at each step their state estimate is optimal, whereas Magill's estimator is only optimal if the true value of the parameter precisely matches one of the elemental filters in the bank.

Later in 1969, Sengbush and Lainiotis [174] proposed a binary method to quantize the parameter space efficiently; the discretization process must be fine enough so that the true parameter value can be accurately estimated, but since computer resources are finite, the quantization must also be sufficiently coarse. Their technique is itself iterative in nature and only requires two quantization levels for each parameter (of the uncertain parameter vector) being estimated.

In 1970, Ackerson and Fu [2] generalized Magill's work when they proposed a method to extend the discrete Kalman-Bucy filter by allowing a nonstationary noise process consisting of a group of Gaussian distributions to drive the filter. They allowed the input and noise process statistics to change (or switch) in discrete "jumps" according to a Markov transition process matrix. Hence the system can be characterized by a model with Markov switching parameters. Whereas Magill's formulation employed multiple models to identify the statistics of a static or unchanging system,

this work used multiple models to characterize a dynamic, time-varying system. Since the optimal algorithm suffers from needing an ever-growing amount of memory³, they proposed a suboptimal finite-memory estimator that assumes that the hypothesis conditional probability is normally distributed when in fact it is a sum of Gaussians that grows exponentially with increasing time.

In 1971, Lainiotis [107] showed how an estimator, from a class of nonlinear adaptive estimators, may be decomposed into a nonadaptive part (the bank of Kalman or Kalman-Bucy filters) and a nonlinear adaptive part which is tasked with identifying the “mode” of the system. A weighted sum of hypothesis conditional probabilities is used to identify the true system mode. He applied this decomposition to the problem of state estimation with non-Gaussian initial state.

In 1973, Moose and Wang [151] proposed modeling the modes or states of the system with a semi-Markov process. In a semi-Markov process, the transitions between states are dictated by the familiar Markov transition matrix; however, the amount of time spent in the current state before switching to the next state is a random variable, i.e., not constant as with a Markov process. They claimed that this modification completely solved the problem of needing increasing computer storage capacity with increasing time. Two years later, Moose [150] applied this formulation to the maneuvering target problem.

In 1974, Fry and Sage [59] employed the hierarchical estimation theory developed by Smith and Sage [179] to reduce the computational requirements of Magill’s method⁴. The hierarchical approach is used to decompose a complex system into

³To characterize the parameter history fully requires K^i hypotheses — and thus K^i elemental filters are required. For example, at time t_1 , we have, for all intents and purposes, the constant parameter case and thus we need just K elemental filters; one for each of the assumed values that the parameter may assume. Then, at time t_2 , we now require K^2 elemental filters since there are now K^2 possible parameter value trajectories, and so on. Consequently, the number of elemental filters, and hence the memory required, grows exponentially when we desire an elemental filter that can exactly match the parameter’s time history.

⁴This paper by Fry and Sage contains an excellent review of Magill’s paper; see also, Maybeck’s second volume for a review of the multiple model adaptive filter [130].

several simple subsystems. The melding of these two techniques enables the application of multiple model methods for hierarchically structured problems that would otherwise require an insurmountable computation load.

In 1976, Lainiotis unified many of the ideas regarding partitioned or MMAE techniques in [108] and presented the idea of cascading controllers with the elemental filters [109]. Thus, multiple model adaptive control (MMAC) was born.

In 1976, Hawkes and Moore [75, 76] reported two important results. They calculated an upper bound for the mean-squared error obtained for a finite parameter set assumption. Secondly, they established some necessary and sufficient conditions for exponential convergence of the Bayesian estimate to the true values in the mean-squared error sense for systems with measurements corrupted by stationary zero-mean Gaussian random processes.

In 1977, a group led by Athans [10] devised the first practical implementation of MMAE and MMAC in a problem that failed to showcase the potential of the multiple model methods because the F-8C aircraft flight controller did not need an adaptive estimator or controller. Nonetheless, hundreds of researchers have contributed many articles and books devoted to estimation (and control) using multiple model methods in the past thirty years; this large volume alone is an indication of its utility and applicability.

In 1978, Chang and Athans [34] proved that if one of the models in the set of K constant-parameter elemental filters exactly matched the real world system, then an MMAE based estimator was optimal. In the event that we don't discretize the parameter space such that one of the K elemental filter models is the truth, then we may say that the MMAE will converge to the closest hypothesized model in the Baram sense [16, 15, 175, 177]. However, no research has yet given us a guaranteed convergence *rate* [79]. Additionally, Chang and Athans proposed an optimal estimator using K^2 elemental filters for the case in which the unknown parameter vector is allowed to vary or switch, specifically, when it follows a Markov process, i.e., the

present parameter vector value depends only on the previous parameter vector value. This estimator is known as a switching parameter algorithm.

In 1979, Tugnait [195] pointed out that Chang and Athans' development [34] was actually suboptimal for the Markov parameter case. Tugnait stated that the Chang and Athans paper approximated the probability density function (PDF) (at time t_i) for each of the K^2 elemental filter residual processes using a single Gaussian PDF. The actual PDF needed to determine the true state optimally for each elemental filter was a Gaussian mixture — a weighted⁵ sum of Gaussian PDFs, K^{i-2} in this case.

In 1980, Tugnait [196] investigated the behavior of a Bayes optimal estimator with unknown continuous parameter vector. Specifically, he studied the convergence properties of a conditional mean estimator in which the unknown parameter is to be determined from an infinite countable set. He also applied his results to a linear time-invariant Gauss-Markov system model.

In 1983, Dasgupta and Westphal [41] extended Hawkes and Moore's [76] convergence results to include systems with unknown biases. They note that in simulations, oftentimes the multiple model estimator preferred the zero-mean elemental filter, hence implementation of non-zero-mean filters should be considered carefully.

In 1984, Blom [22, 23] proposed a computationally efficient algorithm for filtering a system characterized by Markov switching parameters — a problem previously investigated by Ackerson and Fu [2]. The interacting multiple model (IMM) technique reduces the required number of elemental filters in the filter bank through a novel hypothesis merging routine.

Additionally, many researchers have explored MMAE methods to improve the operating characteristics of a system through feedback control. MMAE-based control occurs when the control action is based on a state estimate provided by the MMAE

⁵The weights sum to one so that the mixture retains all the properties of a PDF.

estimator; see for example Stepaniak and Maybeck [186]. On the other hand, MMAC employs a parallel bank of controllers, each matched to a particular filter within the bank of state estimators. As previously noted, Lainiotis [109] was the first to propose this pairing of filters and controllers; a group led by Athans was the first to implement it [10]. The research developed in this dissertation is primarily concerned with MMAE; however, on occasion, we shall discuss MMAE-based control and MMAC to shed additional light on the estimation framework itself.

2.3 Multiple Model Adaptive Estimation Fundamentals

MMAE employs a parallel bank of elemental filters to process noise-corrupted measurements and recursively identify uncertain parameters, estimate states, and compute the residuals between model-based measurement predictions and actual observed measurements. As such, an MMAE algorithm can adapt itself to an uncertain noise environment, perform parameter (or system mode) identification, and compute an accurate state estimate. To accomplish these tasks, the MMAE algorithm processes known inputs and noise-corrupted measurements at discrete times with a set of parallel elemental filters which are developed using a mathematical system model based on a pair of stochastic equations representing the internal state dynamics and measurement processes. Each of the filters in the bank represents a possible mode of the system; each filter is designed using a different hypothesis about the assumed value for the parameters used to describe the structure of the dynamics or measurement models and/or characterize the statistical properties of the dynamics and measurement noise.

The MMAE framework is shown graphically in Figure 2.1. The noise sources are not explicitly labeled and the time dependence has been suppressed in this diagram. The system processes known inputs \mathbf{u} and corrupted measurements \mathbf{z} with a parallel set of K elemental filters F_1, F_2, \dots, F_K based on parameter vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_K$, respectively. Each elemental filter produces a state estimate $\hat{\mathbf{x}}_k$, a

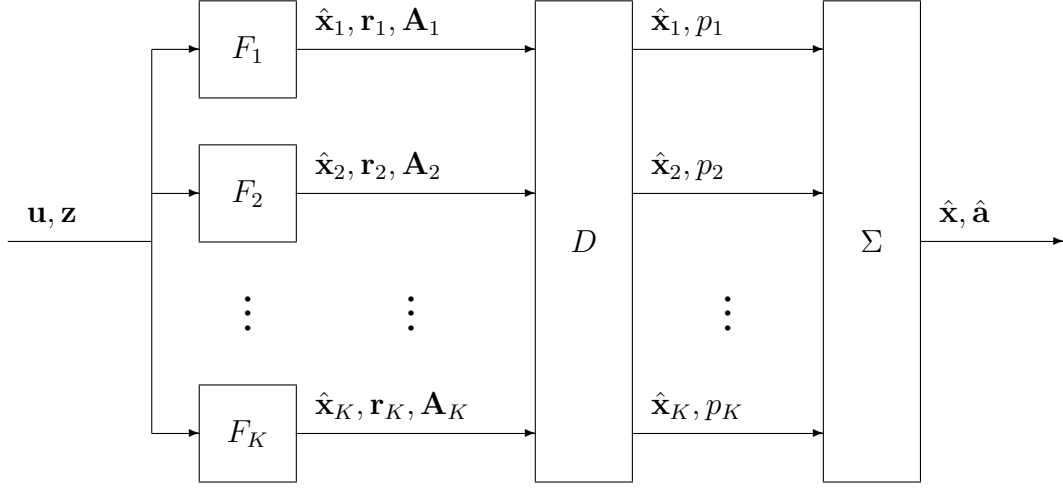


Figure 2.1 Multiple Model Adaptive Estimation

measurement residual \mathbf{r}_k , and a filter-computed residual covariance matrix \mathbf{A}_k . The measurement residuals and covariance matrices are used in block D to compute the probability p_k of the assumed system mode parameter vector \mathbf{a}_k matching the true parameter. In Σ , the probabilities are used in conjunction with the individual state estimates $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_K$ and the known parameters $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_K$ to estimate the system state $\hat{\mathbf{x}}$ and identify the system mode parameter vector $\hat{\mathbf{a}}$. The following sections fill in the details on the mathematical system model, the Kalman filtering algorithm, the filter bank, the parameter and state estimates, and several key assumptions driving the MMAE methodology.

2.3.1 Mathematical System Model. The performance of any model-based algorithm depends heavily on creating an accurate model of the system. The real world rarely, if ever, presents us with a truly linear system, but over limited operating regimes, many of the systems of interest to us can be adequately modeled as linear. Additionally, the disturbances to the system are often well modeled by a vector of additive white Gaussian noise processes. It is often useful to denote both scalar and vector Gaussian stochastic processes using the probability distribution notation [14]: $\mathfrak{N}[\mathbf{x}(t); \boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t)]$, where $\mathfrak{N}[\cdot; \cdot, \cdot]$ is the Gaussian (normal) probability distribution,

$\mathbf{x}(t)$ is the vector stochastic process in question at time t , $\boldsymbol{\mu}(t)$ the mean vector, and $\boldsymbol{\Sigma}(t)$ the covariance matrix.

For this research, we shall assume that our linear system model⁶ has continuous-time dynamics

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t) \mathbf{x}(t) + \mathbf{B}(t) \mathbf{u}(t) + \mathbf{G}(t) \mathbf{w}(t) \quad (2.1)$$

where $\dot{\mathbf{x}}(t) \triangleq d\mathbf{x}(t)/dt$, with measurements available at discrete times

$$\mathbf{z}(t_i) = \mathbf{H}(t_i) \mathbf{x}(t_i) + \mathbf{v}(t_i) \quad (2.2)$$

and is driven by known inputs, $\mathbf{u}(t)$, and independent⁷, zero-mean white⁸ Gaussian noise processes, $\mathbf{w}(\cdot, \cdot)$ and $\mathbf{v}(\cdot, \cdot)$, with known strength $\mathbf{Q}(t)$ and covariance $\mathbf{R}(t_i)$, respectively, and

$$\begin{aligned} \mathbf{x}(t) &= n \times 1 \text{ state vector at time } t \\ \mathbf{F}(t) &= n \times n \text{ system dynamics matrix at time } t \\ \mathbf{B}(t) &= n \times r \text{ input distributor matrix at time } t \\ \mathbf{u}(t) &= r \times 1 \text{ control vector at time } t \\ \mathbf{G}(t) &= n \times s \text{ noise distributor matrix at time } t \\ \mathbf{w}(t) &= s \times 1 \text{ Gaussian noise process vector at time } t \\ \mathbf{z}(t_i) &= m \times 1 \text{ measurement vector at time } t_i \\ \mathbf{H}(t_i) &= m \times n \text{ measurement distributor matrix at time } t_i \\ \mathbf{v}(t_i) &= m \times 1 \text{ Gaussian measurement noise process vector at time } t_i \end{aligned}$$

Additionally, these independent noise processes have covariance kernels of $E \{ \mathbf{w}(t) \mathbf{w}^T(t') \} = \mathbf{Q}(t) \delta(t - t')$ and $E \{ \mathbf{v}(t_i) \mathbf{v}^T(t_j) \} = \mathbf{R}(t_i) \delta_{ij}$. Recall that the

⁶This section is entirely based on Maybeck [129]. Several other books such as McGarty [141] also provide this background in a similar notation.

⁷This assumption is not required but simply makes the presentation easier; for a discussion on how to model correlated noise processes, see Chapter 5 of Maybeck [129] or [181].

⁸A noise process (or a noise sequence) that is independent in time is known as a *white* random process (sequence). While continuous-time white processes don't exist in the real world, this assumption is well justified when the true band-limited noise process frequency bandwidth is much larger than the system bandwidth [129]. Additionally, time-correlated "colored" noise is treated by Maybeck [129] (in Chapter 4), see also [181].

Kronecker delta is defined by [7, 154]

$$\delta_{ij} \triangleq \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (2.3)$$

and the Dirac delta function, $\delta(t)$, is defined as the function that satisfies the following [129, 28]

$$\int_{-\infty}^{\infty} \delta(\tau) d\tau = 1 \text{ and } \delta(\tau) = 0 \text{ for all } \tau \neq 0 \quad (2.4)$$

A more rigorous approach to modeling continuous-time system dynamics⁹ would employ a true differential equation driven by a Brownian motion (or Wiener) process, $\mathbf{b}(t)$, with diffusion $\mathbf{Q}(t)$, versus the more familiar derivative-based Equation (2.1) driven by a zero-mean white Gaussian noise:

$$d\mathbf{x}(t) = [\mathbf{F}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)]dt + \mathbf{G}(t)d\mathbf{b}(t) \quad (2.5)$$

where $\mathbf{b}(t)$ is an $s \times 1$ Brownian motion noise process vector at time t , having a diffusion of $\mathbf{Q}(t)$; the hypothetical derivative of $\mathbf{b}(t)$ would be the $\mathbf{w}(t)$ in Equation (2.1).

Since our algorithm will be implemented on a digital computer, we require an equivalent discrete-time model for the system dynamics¹⁰. We shall begin with Equation (2.5) to create a stochastic difference equation; we only report the results here — see, for example, Maybeck [129], for the proper procedure, or our development in Section 3.4 for a more general case. Thus our mathematical model of the system dynamics and measurement process becomes

$$\mathbf{x}(t_{i+1}) = \mathbf{\Phi}(t_{i+1}, t_i)\mathbf{x}(t_i) + \mathbf{B}_d(t_i)\mathbf{u}(t_i) + \mathbf{G}_d(t_i)\mathbf{w}_d(t_i) \quad (2.6)$$

$$\mathbf{z}(t_i) = \mathbf{H}(t_i)\mathbf{x}(t_i) + \mathbf{v}(t_i) \quad (2.7)$$

⁹The measurement model of Equation (2.2) is unchanged.

¹⁰Here we implicitly assume that the control input $\mathbf{u}(t)$ is a piece-wise constant function such that $\mathbf{u}(t) = \mathbf{u}(t_i)$ for time $t_i \leq t < t_{i+1}$.

where

$$\begin{aligned}
\mathbf{x}(t_i) &= n \times 1 \text{ state process vector at time } t_i \\
\Phi(t_{i+1}, t_i) &= n \times n \text{ state transition matrix from time } t_i \text{ to time } t_{i+1} \\
\mathbf{B}_d(t_i) &= n \times r \text{ discrete-time input distributor matrix at time } t_i \\
\mathbf{u}(t_i) &= r \times 1 \text{ discrete-time control vector at time } t_i \\
\mathbf{G}_d(t_i) &= n \times s \text{ discrete-time noise distributor matrix at time } t_i \\
\mathbf{w}_d(t_i) &= s \times 1 \text{ discrete-time white Gaussian noise process vector at time } t_i \\
\mathbf{z}(t_i) &= m \times 1 \text{ measurement process vector at time } t_i \\
\mathbf{H}(t_i) &= m \times n \text{ measurement distributor matrix at time } t_i \\
\mathbf{v}(t_i) &= m \times 1 \text{ white Gaussian measurement noise process vector at time } t_i
\end{aligned}$$

and the discrete-time noise distributor matrix is chosen without loss of generality to be an $n \times n$ identity matrix: $\mathbf{G}_d(t_i) = \mathbf{I}$.

Since we began with a continuous-time dynamics model, the state transition matrix must satisfy the following differential equation with initial condition

$$\begin{aligned}
d\Phi(t, t_0)/dt &= \mathbf{F}(t) \Phi(t, t_0) \\
\Phi(t_0, t_0) &= \mathbf{I}
\end{aligned} \tag{2.8}$$

The state transition matrix has several important properties:

1. $\Phi(t, t')$ is uniquely defined for all times t and t' in $[0, \infty)$.
2. Semi-group property: $\Phi(t'', t) = \Phi(t'', t') \Phi(t', t)$ for any times $t, t', t'' \in [0, \infty)$ with $t \leq t' \leq t''$.
3. Semi-group property, special case: $\Phi(t', t) \Phi(t, t') = \mathbf{I}$ for any times $t, t' \in [0, \infty)$ with $t \leq t'$.
4. Nonsingular¹¹: $\Phi^{-1}(t, t_0) = \Phi(t_0, t)$ for any times $t, t_0 \in [0, \infty)$ with $t \geq t_0$.

When \mathbf{F} is time-invariant, i.e., a *constant matrix*, then the state transition matrix becomes a function of the difference of the time arguments and is explicitly represented

¹¹The state transition matrix is guaranteed to be nonsingular when we begin with a continuous-time system description; this is not necessarily so for naturally discrete-time systems.

as the matrix exponential

$$\Phi(t, t_0) = \Phi(t - t_0) = \exp\{\mathbf{F}(t - t_0)\} \quad (2.9)$$

The equivalent discrete-time input distributor matrix, $\mathbf{B}_d(t_i)$, is found by integrating the continuous-time input distributor matrix weighted by the state transition matrix over one sample period so that

$$\mathbf{B}_d(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, \tau) \mathbf{B}(\tau) d\tau \quad (2.10)$$

The covariance kernel for the zero-mean white¹² Gaussian dynamics noise sequence is given by

$$E \{ \mathbf{w}_d(t_i) \mathbf{w}_d^T(t_j) \} = \begin{cases} \mathbf{Q}_d(t_i), & t_i = t_j \\ \mathbf{0}, & t_i \neq t_j \end{cases} \quad (2.11)$$

where the continuous-time noise strength $\mathbf{Q}(t)$ is used to determine the positive semi-definite discrete-time noise covariance matrix, $\mathbf{Q}_d(t_i)$, expressed as

$$\mathbf{Q}_d(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, \tau) \mathbf{G}(\tau) \mathbf{Q}(\tau) \mathbf{G}^T(\tau) \Phi^T(t_{i+1}, \tau) d\tau \quad (2.12)$$

The covariance kernel for the zero-mean white Gaussian measurement noise sequence is given by

$$E \{ \mathbf{v}(t_i) \mathbf{v}^T(t_j) \} = \begin{cases} \mathbf{R}(t_i), & t_i = t_j \\ \mathbf{0}, & t_i \neq t_j \end{cases} \quad (2.13)$$

where $\mathbf{R}(t_i)$ is assumed to be positive definite. The discrete-time noise processes $\mathbf{w}_d(\cdot, \cdot)$ and $\mathbf{v}(\cdot, \cdot)$ are assumed to be independent, random processes. Additionally, the initial state condition $\mathbf{x}(t_0)$ is not known precisely; it will be modeled as a Gaussian random vector, independent of both noise processes $\mathbf{w}_d(\cdot, \cdot)$ and $\mathbf{v}(\cdot, \cdot)$,

¹²A white process is independent in time and thus has a zero (matrix) covariance kernel for the $t_i \neq t_j$ case; see the second line of Equations (2.11) and (2.13).

with a mean and covariance of

$$E\{\mathbf{x}(t_0)\} = \hat{\mathbf{x}}_0 \quad (2.14)$$

$$E\{[\mathbf{x}(t_0) - \hat{\mathbf{x}}_0][\mathbf{x}(t_0) - \hat{\mathbf{x}}_0]^T\} = \mathbf{P}_0 \quad (2.15)$$

respectively, where \mathbf{P}_0 is positive semi-definite matrix.

For naturally discrete-time systems, the dynamics model could be described more generally with

$$\begin{aligned} \mathbf{G}_d(t_i) &= n \times s \text{ noise distributor matrix at time } t_i \\ \mathbf{w}_d(t_i) &= s \times 1 \text{ noise process vector at time } t_i \end{aligned}$$

where the noise process $\mathbf{w}_d(\cdot, \cdot)$ is zero-mean white Gaussian noise process vector at time t_i with positive semi-definite $s \times s$ covariance matrix $\mathbf{Q}_d(t_i)$. *However*, all of the equations in this section and those that follow are invariant with respect to the dimensions of \mathbf{G}_d and \mathbf{w}_d .

The time-varying model has been used extensively for state estimation problems, but for identifying parameter variation over a set of constant (or slowly time-varying) parameters, we will generally employ the following *time-invariant* system model equations:

$$\mathbf{x}(t_i) = \mathbf{\Phi} \mathbf{x}(t_{i-1}) + \mathbf{B}_d \mathbf{u}(t_{i-1}) + \mathbf{w}_d(t_{i-1}) \quad (2.16)$$

$$\mathbf{z}(t_i) = \mathbf{H} \mathbf{x}(t_i) + \mathbf{v}(t_i), \quad (2.17)$$

where the time-invariance property gives $\mathbf{B}_d(t_{i-1}) = \mathbf{B}_d$ and $\mathbf{H}(t_i) = \mathbf{H}$, uniform spacing between time samples yields a state transition matrix independent of time, i.e., $\mathbf{\Phi}(t_i, t_{i-1}) = \mathbf{\Phi}$, and stationary noises processes result in: $\mathbf{Q}_d(t_i) = \mathbf{Q}_d$ and $\mathbf{R}(t_i) = \mathbf{R}$ for all time t_i . In this time-invariant case, these zero-mean white Gaussian noise processes $\mathbf{w}_d(\cdot, \cdot)$ and $\mathbf{v}(\cdot, \cdot)$ can be thought of as sequences of independent, identically distributed (IID) random variables and denoted by $\mathbf{w}_d(t_i) \sim \mathfrak{N}(\mathbf{0}, \mathbf{Q}_d)$ and $\mathbf{v}(t_i) \sim \mathfrak{N}(\mathbf{0}, \mathbf{R})$, where the \mathfrak{N} indicates that the random vector process has a

normal or Gaussian distribution [159, 170]. Furthermore, a derivation of the steady-state Kalman filter is based on the time-invariant system model [129] .

For some problems of interest, one or both of the noise sources can be discrete space-time point processes versus continuous-time (for the dynamics) or discrete-time (for the measurement and/or dynamics) Gaussian processes. In order to preserve the Markov nature of the state estimate, the noise process must be an independent increment process, i.e., the noise processes for nonoverlapping (disjoint) periods of time are independent — see Definition 60 in Chapter III. Thus, the Gaussian property assumed for the driving noise during the development of the Kalman filter is not a necessary condition to derive *an* optimal filter, it is merely sufficient since the white Gaussian process is an independent increment process. If the dynamics noise \mathbf{w}_d and/or the measurement noise \mathbf{v} is a generalized Poisson point process, we can still derive an optimal filter. The Snyder filter assumes that the measurement noise process is a Poisson point process [180]. The Snyder filter was employed Meer [142] and others [218, 93, 74] in an MMAE structure to control the pointing and tracking of particle beams. This will not be pursued further herein.

Some useful notation for describing the stochastic measurement history and its realization are defined as:

$$\mathbf{Z}(t_i) \triangleq \begin{bmatrix} \mathbf{z}(t_1) \\ \mathbf{z}(t_2) \\ \vdots \\ \mathbf{z}(t_i) \end{bmatrix} \quad \text{and} \quad \mathbf{Z}_i \triangleq \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_i \end{bmatrix} \quad (2.18)$$

respectively, where \mathbf{z}_i is a convenient notation for $\mathbf{z}(t_i)$, a specific realization of the random vector $\mathbf{z}(t_i)$. Note that these vectors “grow” over time.

2.3.2 Kalman Filtering. An optimal solution to the estimation problem discussed above in Equations (2.6) through (2.15) was given by Kalman in his 1960

paper “A New Approach to Linear Filtering and Prediction Problems” [95]¹³. The following year, Kalman and Bucy presented their solution for the continuous-time problem [96]. A recent derivation of Kalman’s results were presented by Catlin [33] to the mathematical community in his 1989 book: *Estimation, Control, and the Discrete Kalman Filter* as a “beautiful illustration of functional analysis in action” in which the projection theorem in a Hilbert space plays a central role¹⁴. Recall that the intent of this research is to explore the utility of employing *multiple* models in a parallel structure in order to improve state estimation (or to identify the system mode parameters themselves); hence we shall only develop those concepts directly related to the structure. We will follow Maybeck’s treatment of the Kalman filter — an optimal¹⁵ recursive data processing algorithm [129]. The Kalman filter recursively generates the optimal state estimate to the problem posed above with a two stage process: first, it predicts the state for time t_i using only the dynamics model and the measurements up through time t_{i-1} and then corrects or updates the estimate with noise-corrupted measurements¹⁶.

According to the Bayesian viewpoint espoused by Maybeck [129], the sampled-data Kalman filter algorithm consists of (initializing and then) recursively propagat-

¹³Kalman’s original paper has been republished in many collections, such as the Kalman filtering collection edited by Sorenson [183] and another collection edited by Başar that focuses on control theory [12].

¹⁴The infinite-dimensional sampled-data Kalman filter we derive in Chapter 3 is a more general beautiful illustration of functional analysis in action.

¹⁵The Kalman filter produces *the* optimal state estimate, $\hat{\mathbf{x}}(t_i^+)$, for a stochastic linear system driven by zero-mean, white Gaussian noise processes with known covariances [129]! In the Bayesian sense, $\hat{\mathbf{x}}(t_i^+)$ is the optimal state estimate because it is the mean, median, and mode of the Gaussian conditional PDF $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}(\boldsymbol{\xi}|\mathbf{Z}_i)$. $\hat{\mathbf{x}}(t_i^+)$ minimizes the mean-squared error (MSE) and the symmetric cost function by virtue of being the conditional mean of the Gaussian conditional PDF. $\hat{\mathbf{x}}(t_i^+)$ is the maximum *a posteriori* (MAP) state estimate and when there is no initial state information, i.e., $\mathbf{P}_0^{-1} = \mathbf{0}$, then $\hat{\mathbf{x}}(t_i^+)$ is also the maximum likelihood (ML) estimate. When the noises are nonGaussian, the Kalman filter estimate is the optimal *linear* estimator: it is the linear minimum variance unbiased (MVU) estimate; thus, we’re saying that a nonlinear estimator may do better. And finally, as Kalman originally posed [95], $\hat{\mathbf{x}}(t_i^+)$ is the orthogonal projection of the true state $\mathbf{x}(t_i)$ onto the subspace spanned by the random measurement history $\mathbf{Z}(t_i)$, i.e., $\hat{\mathbf{x}}(t_i^+)$ satisfies the projection theorem [122] and is thus the optimal estimate of $\mathbf{x}(t_i)$ given measurements $\mathbf{Z}(t_i)$.

¹⁶For nonlinear models, the extended Kalman filter (EKF) is an appropriate tool; see Maybeck [130] or Sworder and Boyd [193] for information on the EKF and other nonlinear filters.

ing and updating the state conditional PDF. The **sampled-data Kalman filter algorithm** is [129]:

1. *Initialize the Gaussian PDF, $f_{\mathbf{x}(t_0)}(\boldsymbol{\xi})$.* The initial state is modeled by a Gaussian random vector with mean and covariance given by the initial state estimate $\hat{\mathbf{x}}(t_0)$ and initial covariance estimate $\mathbf{P}(t_0)$:

$$\hat{\mathbf{x}}(t_0) \triangleq E\{\mathbf{x}(t_0)\} = \hat{\mathbf{x}}_0 \quad (2.19)$$

and

$$\mathbf{P}(t_0) \triangleq E\{[\mathbf{x}(t_0) - \hat{\mathbf{x}}_0][\mathbf{x}(t_0) - \hat{\mathbf{x}}_0]^T\} = \mathbf{P}_0 \quad (2.20)$$

2. *Propagate the Gaussian conditional PDF.* The propagation is entirely based on the known internal dynamics model conditioned on the observed measurements. This stage predicts the state estimate at time t_i ¹⁷ given the optimal estimate at time t_{i-1} :

$$\begin{aligned} \hat{\mathbf{x}}(t_i^-) &\triangleq E\{\mathbf{x}(t_i) | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\} \\ &= \boldsymbol{\Phi}(t_i, t_{i-1}) \hat{\mathbf{x}}(t_{i-1}^+) + \mathbf{B}_d(t_{i-1}) \mathbf{u}(t_{i-1}) \end{aligned} \quad (2.21)$$

and

$$\begin{aligned} \mathbf{P}(t_i^-) &\triangleq E\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)][\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)]^T | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\} \\ &= \boldsymbol{\Phi}(t_i, t_{i-1}) \mathbf{P}(t_{i-1}^+) \boldsymbol{\Phi}^T(t_i, t_{i-1}) + \mathbf{G}_d(t_{i-1}) \mathbf{Q}_d(t_{i-1}) \mathbf{G}_d^T(t_{i-1}) \end{aligned} \quad (2.22)$$

where the expectation is taken with respect to conditional PDF $f_{\mathbf{x}(t_i) | \mathbf{Z}(t_{i-1})}(\boldsymbol{\xi} | \mathbf{Z}_{i-1})$.

3. *Update the Gaussian conditional PDF.* Update the state and covariance estimates at time t_i with the latest measurement $\mathbf{z}(t_i, \omega_j) = \mathbf{z}_i$ to produce $\hat{\mathbf{x}}(t_i^+)$

¹⁷Time t_i^- represents the time t_i just *prior to* measurement update; some authors [170] use the notation $t_i | t_{i-1}$ and while others [13] often use only the indices, i.e., $i | i - 1$. Time t_i^+ represents the time t_i just *after* measurement update; some authors write this as $t_i | t_i$ or $i | i$.

and $\mathbf{P}(t_i^+)$. The filter-computed residual covariance $\mathbf{A}(t_i)$ and Kalman gain $\mathbf{K}(t_i)$ are computed first as:

$$\mathbf{A}(t_i) \equiv \mathbf{H}(t_i) \mathbf{P}(t_i^-) \mathbf{H}^T(t_i) + \mathbf{R}(t_i) \quad (2.23)$$

and

$$\mathbf{K}(t_i) \equiv \mathbf{P}(t_i^-) \mathbf{H}^T(t_i) \mathbf{A}^{-1}(t_i) \quad (2.24)$$

Using the conditional PDF, $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}(\boldsymbol{\xi}|\mathbf{Z}_i)$, we can compute the optimal state estimate¹⁸ given by

$$\begin{aligned} \hat{\mathbf{x}}(t_i^+) &\triangleq E \{ \mathbf{x}(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i \} \\ &= \hat{\mathbf{x}}(t_i^-) + \mathbf{K}(t_i) \mathbf{r}(t_i) \end{aligned} \quad (2.25)$$

where the Kalman filter residual $\mathbf{r}(t_i)$ is

$$\mathbf{r}(t_i) \equiv \mathbf{z}_i - \mathbf{H}(t_i) \hat{\mathbf{x}}(t_i^-) \quad (2.26)$$

where $\mathbf{H}(t_i) \hat{\mathbf{x}}(t_i^-)$ is the predicted measurement sometimes denoted $\hat{\mathbf{z}}(t_i^-)$, and the updated error covariance is

$$\begin{aligned} \mathbf{P}(t_i^+) &\triangleq E \{ [\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)] [\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)]^T | \mathbf{Z}(t_i) = \mathbf{Z}_i \} \\ &= \mathbf{P}(t_i^-) - \mathbf{K}(t_i) \mathbf{H}(t_i) \mathbf{P}(t_i^-) \end{aligned} \quad (2.27)$$

4. Return to step 2

2.3.3 Filter Bank. A natural way to extend the concept of state estimation using a single filter into the realm of joint state and parameter estimation is to employ

¹⁸The state estimate $\hat{\mathbf{x}}(t_i^+)$ is the sum of a *prediction* $\hat{\mathbf{x}}(t_i^-)$, which is a sufficient statistic for the state $\mathbf{x}(t_i)$ given $\mathbf{Z}(t_{i-1})$ [100], and a *correction* term $\mathbf{K}(t_i)\mathbf{r}(t_i)$ which represents the “new information” provided by the current measurement.

a parallel bank of filters, i.e., a *filter bank*. Observe that the conditional PDF for the state (at measurement update) can be interpreted as a *marginal* conditional PDF computed from the *joint* conditional PDF, i.e.,

$$f_{\mathbf{x}(t_i)|\mathbf{z}(t_i)}(\boldsymbol{\xi}|\mathbf{Z}_i) = \int_{-\infty}^{\infty} f_{\mathbf{x}(t_i),\mathbf{a}(t_i)|\mathbf{z}(t_i)}(\boldsymbol{\xi}, \boldsymbol{\alpha}|\mathbf{Z}_i) d\boldsymbol{\alpha} \quad (2.28)$$

where the system mode $\mathbf{a}(t_i)$ is a random vector and $\boldsymbol{\alpha}$ represents a particular mode of the system. While the structure of an elemental filter is designed using the same mathematical system model based on a pair of stochastic equations representing the internal state dynamics and measurement processes, each mode of the system is characterized by a unique parameter vector and described probabilistically by a Gaussian PDF: $f_{\mathbf{x}(t_i)|\mathbf{a}(t_i),\mathbf{z}(t_i)}(\boldsymbol{\xi}|\boldsymbol{\alpha}, \mathbf{Z}_i)$, with mean $\hat{\mathbf{x}}(t_i^+)$ and covariance $\mathbf{P}(t_i^+)$ as computed by a Kalman filter based on the parameter value $\mathbf{a}(t_i) = \boldsymbol{\alpha}$. From the Bayesian point of view, we are motivated to pursue the joint state and parameter vector conditional PDF:

$$f_{\mathbf{x}(t_i),\mathbf{a}(t_i)|\mathbf{z}(t_i)}(\boldsymbol{\xi}, \boldsymbol{\alpha}|\mathbf{Z}_i) = f_{\mathbf{x}(t_i)|\mathbf{a}(t_i),\mathbf{z}(t_i)}(\boldsymbol{\xi}|\boldsymbol{\alpha}, \mathbf{Z}_i) f_{\mathbf{a}(t_i)|\mathbf{z}(t_i)}(\boldsymbol{\alpha}|\mathbf{Z}_i) \quad (2.29)$$

since it features all of the variables that we are interested in estimating, given all of the available measurements. The second term of Equation (2.29) is

$$f_{\mathbf{a}(t_i)|\mathbf{z}(t_i)}(\boldsymbol{\alpha}|\mathbf{Z}_i) = \sum_{k=1}^K p_k(t_i) \delta(\boldsymbol{\alpha} - \mathbf{a}_k) \quad (2.30)$$

with $p_k(t_i)$ defined by

$$p_k(t_i) \triangleq \text{pr}\{\mathbf{a}(t_i) = \mathbf{a}_k | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \quad (2.31)$$

where \mathbf{a}_k is the k th system mode parameter vector.

Once Equation (2.29) is so established, we can write (via marginal PDFs):

$$f_{\mathbf{x}(t_i)|\mathbf{z}(t_i)}(\boldsymbol{\xi}|\mathbf{Z}_i) = \int_{-\infty}^{\infty} f_{\mathbf{x}(t_i),\mathbf{a}(t_i)|\mathbf{z}(t_i)}(\boldsymbol{\xi}, \boldsymbol{\alpha}|\mathbf{Z}_i) d\boldsymbol{\alpha} \quad (2.32)$$

Then applying the sifting property of the Dirac delta function, $\delta(\cdot)$, yields

$$f_{\mathbf{x}(t_i)|\mathbf{z}(t_i)}(\boldsymbol{\xi}|\mathbf{Z}_i) = \sum_{k=1}^K f_{\mathbf{x}(t_i)|\mathbf{a}(t_i),\mathbf{z}(t_i)}(\boldsymbol{\xi}|\mathbf{a}_k, \mathbf{Z}_i) \text{pr}\{\mathbf{a}(t_i) = \mathbf{a}_k | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \quad (2.33)$$

Hence, the development of the single filter as a lumped expression of a (truly distributed) system is naturally represented by a filter bank via the total probability theorem¹⁹.

We will continue to consider and develop these ideas in the sections that follow in terms of the concept of the uncertain parameter and the parameter space discretization process — a process which can be viewed using the total probability theorem. Finally, the state and parameter estimates generated by the MMAE are given and then the assumptions underlying the (static) multiple model methods are reviewed. Later in the chapter, in Sections 2.4.7, we introduce and briefly discuss several dynamic multiple model techniques.

2.3.3.1 Parameter Vector. The first step in building a filter bank is to identify the parameter vectors which we use to represent the system modes²⁰. From the discussion in Section 2.3.1, we know that the elements of the matrices: $\boldsymbol{\Phi}(t_i, t_{i-1})$, $\mathbf{B}_d(t_i)$, $\mathbf{G}_d(t_i)$, $\mathbf{H}(t_i)$, $\mathbf{Q}_d(t_i)$, $\mathbf{R}(t_i)$ describe the structure and/or characterize the statistics of the dynamics and measurement models given, respectively, in Equations (2.6) and (2.7). The elements of these matrices can be functions of a set of quantities that are called the *parameters*; each scalar parameter can affect one or more elements of these matrices. Together the parameters, equations, and

¹⁹Compare this observation with the virtual filter bank discussed in Section 2.5.

²⁰In actuality, the parameter vector represents the portion of the system model which varies and consequently gives rise to the different system modes.

other assumptions and comments define the mathematical system model. Our goal is to estimate a small subset of these parameters (which are assumed to be constant over time for a static filter bank or allowed to vary slowly²¹ with time for a dynamic or moving bank of filters) at each point in time t_i ; this estimation process is called *parameter identification*²².

Specifically, the parameter vector $\mathbf{a}(t_i)$ represents uncertainty in any of the elements of $\Phi(t_i, t_{i-1})$, $\mathbf{B}_d(t_i)$, $\mathbf{G}_d(t_i)$, $\mathbf{H}(t_i)$, $\mathbf{Q}_d(t_i)$, or $\mathbf{R}(t_i)$. It is important to note that the uncertainty in $\Phi(t_i, t_{i-1})$, $\mathbf{B}_d(t_i)$, $\mathbf{G}_d(t_i)$, or $\mathbf{Q}_d(t_i)$ may be due to uncertainty in the continuous-time dynamics structure $\mathbf{F}(t)$. Uncertainties in the plant noise distributor $\mathbf{G}(t)$ or $\mathbf{G}_d(t_i)$ are treated equivalently as uncertainties in $\mathbf{Q}_d(t_i)$ and oftentimes, we roll the uncertainties in $\mathbf{H}(t_i)$ into either $\Phi(t_i, t_{i-1})$ or $\mathbf{B}_d(t_i)$ by an alternative choice of state variables, since we often cannot isolate both at the same time.

This subset of uncertain parameters is modeled as a slowly varying discrete random process²³ and is denoted by $\mathbf{a}(t_i)$. Oftentimes the choice is obvious, but when it is not, an empirical study is conducted on the entire list of parameters to determine the J parameters most crucial for the task at hand; this analysis often depends on whether we are most interested in identifying the parameter in force, improving the state estimate, or enhancing control action [175]. The random parameter vector (representing the system mode that may vary with time) and its realization are

²¹Slowly as compared to the dominant time constants of the system or measurement process.

²²For a comprehensive explication of the standard methods see, for example, the fine texts by Sorenson [182], Sage and Melsa [167], and Ljung [121]

²³Much of the literature, beginning with Magill [125], takes the approach of calling the parameter vector a deterministic quantity that is simply a collection of unknown constants and uses an analysis of the residuals to determine which filter or linear combination of filters best estimates this parameter value.

denoted, respectively, by

$$\mathbf{a}(t_i) \triangleq \begin{bmatrix} \mathbf{a}_1(t_i) \\ \mathbf{a}_2(t_i) \\ \vdots \\ \mathbf{a}_J(t_i) \end{bmatrix} \quad \text{and} \quad \mathbf{a}(t_i) \triangleq \begin{bmatrix} a_1(t_i) \\ a_2(t_i) \\ \vdots \\ a_J(t_i) \end{bmatrix} \quad (2.34)$$

Each element of the realization is defined on a subset of the real number line

$$a_j \in \mathbb{A}_j \subset \mathbb{R}, \forall j = 1, 2, \dots, J \quad (2.35)$$

and the entire *parameter space* is denoted by a product set

$$\mathbb{A} \triangleq \mathbb{A}_1 \times \mathbb{A}_2 \times \dots \times \mathbb{A}_J \subset \mathbb{R}^J \quad (2.36)$$

where [154]

$$\mathbb{A}_1 \times \mathbb{A}_2 \times \dots \times \mathbb{A}_J \triangleq \{(a_1, a_2, \dots, a_J) | a_j \in \mathbb{A}_j \forall j = 1, 2, \dots, J\} \quad (2.37)$$

These subsets, \mathbb{A}_j for every $j = 1, 2, \dots, J$, on the real line may be discrete, continuous, or mixed. Within this work, we will sometimes use the terms “parameter vector”, “mode”, and “model” interchangeably even though the parameter vector only refers to part of the model used to represent the difference between the system modes.

An example will help to clarify this notation. Let $J = 2$ and thus $\mathbf{a} = [a_1 \ a_2]^T$. Then let $a_1 \in \mathbb{A}_1 = [0, \infty)$ represent the unknown nonnegative scalar multiplier used to specify the dynamics noise strength, $\mathbf{Q} = a_1 \mathbf{I}$, where \mathbf{I} is an identity matrix of the appropriate size. Next, let $a_2 \in \mathbb{A}_2 = \{0, 1\}$ be an important parameter in the dynamics matrix, \mathbf{F} , that is either zero or one. If we were to conduct an exper-

iment like this, we would be performing system identification in an uncertain noise environment. Finally, we see that parameter space $\mathbb{A} = \mathbb{A}_1 \times \mathbb{A}_2 = [0, \infty) \times \{0, 1\}$ is indeed a subset of \mathbb{R}^2 . Since the MMAE fundamentally assumes that a parameter may only assume values from among a finite set, we will need to approximate the nonnegative real line $[0, \infty)$ with a discrete set of points versus the continuous set of points currently assumed. How to choose the “best” set of points is still an active area of research.

2.3.3.2 Parameter Space Discretization.

The earliest attempt to sample or discretize the admissible set of parameter values was accomplished by Sengbush and Lainiotis [174]; they proposed two algorithms for a binary quantization of the admissible set. A decade later, Lamb and Westphal [110] used a simplex method of nonlinear programming to direct the discretization process.

The process of “choosing” K points in the parameter space is often called *discretization*; the collection of the K points is called the parameter set²⁴. The goal of parameter discretization is to represent the parameter space accurately with a small set of discrete points in order to reduce the computational burden *and* to increase the distinguishability of the elemental filters in the filter bank. The success of the MMAE depends on the distinguishability of the models used in the bank of elemental filters. To determine which parameter value to use, there must be appreciable differences between the characteristics of the residuals for the “correct” model versus the other, mismatched, filters. Additionally, when Kalman filters are used in the bank, conservative tuning should be avoided to prevent the residuals from becoming too close together and affecting the discrimination of the algorithm; this effect will be discussed in more depth later. In the limit as the residuals become indistinguishable, the adaptation process is totally incapacitated. For fast and reliable parameter

²⁴Some authors refer to this entire process as defining the *model set*, i.e., defining which models to use in the bank of filters, see for example [120, 117, 119]. Note that defining the model set is more general since it also includes filter bank composition.

identification, assuming one of the hypothesized filters models is based on the true parameter value, the residuals should be as distinct as possible [124, 130].

The set \mathbb{A} is the admissible set of parameter values, called *points*, that the parameter vector may assume. This admissible set is normally a subset of J -dimensional Euclidean space and is commonly called the *parameter space*. In order to implement the multiple model algorithm, the designer must choose (in some intelligent fashion) a subset of points from \mathbb{A} in order to represent the parameter space²⁵.

Continuing the example of the previous section, we choose a maximum practical value for the set \mathbb{A}_1 to be 10. Thus we now have²⁶ the closed interval $\mathbb{A}_1 = [0, 10]$; we shall discretize it into the set $\{0, 5, 10\}$. After discretizing \mathbb{A}_1 we have²⁷ $\mathbb{A} = \{0, 5, 10\} \times \{0, 1\} = \{(0, 0), (5, 0), (10, 0), (0, 1), (5, 1), (10, 1)\}$. We now have six points in \mathbb{A} from which to choose, thus the number of elemental filters, K , is equal to six. Additionally, we shall assume that all six of these points represent legitimate parameter values with which we can design a filter. Since our discretization of \mathbb{A}_1 was completely arbitrary, it is possible that an alternate discretization of the half-line $\mathbb{A}_1 = [0, \infty)$ would yield better results in terms of improved state or parameter estimation.

The simplest (and most likely the least effective [177]) approach to discretizing the parameter space is to divide the domain uniformly for each of the J parameters in the parameter vector into $N_j - 1$ intervals, where N_j can be different for each of the J parameters, and then design a filter at each boundary point [53, 98]. (Note that this is the method that we used to discretize $\mathbb{A}_1 = [0, 10]$ in the example a couple of paragraphs back, where $N_1 = 3$ gave rise to two intervals and the set $\{0, 5, 10\}$.)

²⁵Magill [125] assumed that the parameter space was populated with a finite set of known values.

²⁶Note that we have used the same notation for the set \mathbb{A}_1 both before and after discretization in the same spirit as when a computer program assigns a new value to an existing variable.

²⁷As with the set \mathbb{A}_1 , we have redefined the product space \mathbb{A} after discretization process to be \mathbb{A} .

A slightly more effective method would divide the domain uniformly for each of the J parameters in the parameter vector into N_j intervals and then design a filter at the center of each interval [177]; for this case, we would get $\mathbf{A}_1 = \{1\frac{2}{3}, 5, 8\frac{1}{3}\}$. In either case, we construct $N_1 N_2 = K$ filters. For some parameters, it might make sense to space the intervals logarithmically²⁸ [81, 136]. In theory and in practice, the parameter space does not have to be convex, thus K is the maximum number of distinct filters that can be constructed from the product set. In other words, one or more of the defining J -tuple points may, in fact, be invalid. These simple methods are accomplished off-line and the number of filters is held constant.

A better *ad hoc* sampling scheme would include some measure of performance to control the discretization process and would (most likely) result in a nonuniformly sampled space. We could begin by designing the first Kalman filter in the bank at some nominal point. Then, while monitoring the estimation accuracy (or some function of the residuals²⁹ or estimation errors), vary one parameter in one direction at a time. Choose the points which only allow the accuracy to degrade by some set amount [113, 128, 114, 49, 50].

Recently, Erickson [49, 50] discretized his parameter set by monitoring an information distance measure [16] as he varied a single parameter. This so-called Baram distance is basically the likelihood quotient of the Gaussian conditional PDF, which will be discussed and defined in Section 2.3.3.3, Equation (2.44). Since he had seven measurements, he expected the true value of the likelihood quotient to be seven for a properly tuned Kalman filter. He found that, by choosing his parameter points such that the likelihood quotient increased to the same value, which in this

²⁸This simple scheme is utilized in this research even though it may not be optimal for our problem.

²⁹The residual, which was defined mathematically in Equation (2.26), is simply the difference of the predicted and observed measurements and thus contains information on how well a filter model matches up against the true system. An appropriate and hence very common function of the measurement residual vector is the likelihood quotient defined in Equation (2.46): $\mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)$.

case was 14, gave a viable parameter variation in all directions. He also invoked the idea of foveal versus peripheral regions³⁰ for filter bank discretization.

Sheldon [175, 176, 177] and Lund [123, 124] have both contributed to understanding parameter set creation via the discretization process. Sheldon’s main contribution was an optimal discretization procedure that allowed the designer to focus on state or parameter estimation or control regulation using a user specified cost functional. A weighting matrix allows the designer to tailor the weight placed on each state, parameter, or control action. Lund proposed an online algorithm that would maintain the distinguishability (by lowering the dynamics noise strength \mathbf{Q} or Kalman filter gain \mathbf{K}) of the elemental filters as environment changes necessitated modifying the existing filter bank. Lund’s work was extended for the discrete-time case by Miller [149] and Vasquez [198].

Vasquez [198] and Miller [149] modified Sheldon’s “static” algorithm to provide on-line discretization of an adaptive MMAE bank; we shall call this *dynamic* discretization as the filters are allowed to be based upon piecewise constant parameter values rather than constant ones, to respond better to a nonstationary environment. Previous approaches required that the moving-bank algorithm store predetermined discretizations [130, 132, 69, 68, 172]. This dynamic algorithm can run in real-time and the optimization assumes a finite horizon (rather than an infinite horizon and steady state values) in the computation of the discrete parameter values. Discretizations do not have to be predetermined although filters may be pre-computed and stored for speed enhancement.

A few observations regarding the discretization process:

1. Since we are unlikely to have a model that exactly matches the real world in the bank, we can still attain good performance if we interpolate between the

³⁰An analogy to the foveal and peripheral regions of human eyesight, where the foveal high-resolution vision is concentrated at the center of the field of view and peripheral low-resolution vision is more sensitive to light changes and covers the remaining portion of the field of view.

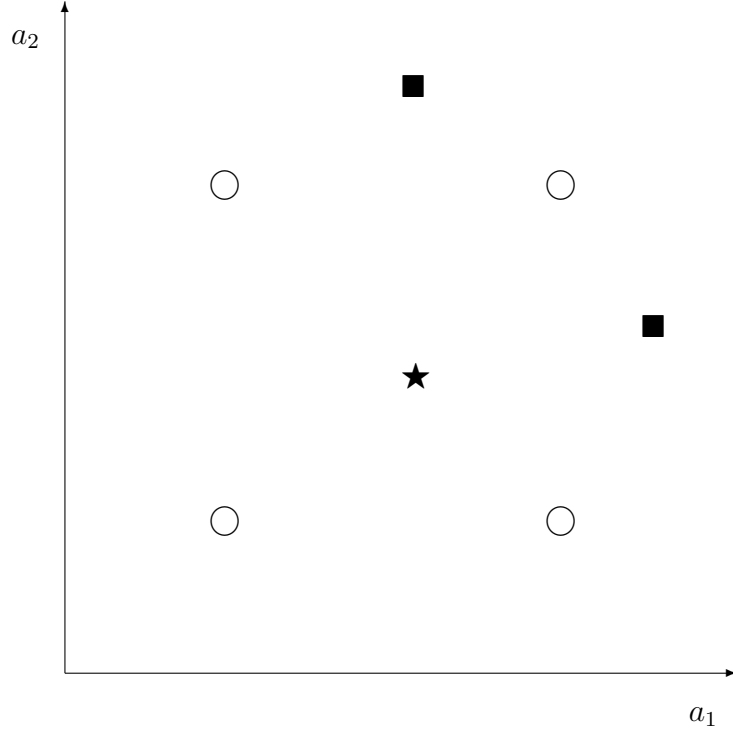


Figure 2.2 Surround the true parameter. Legend: ○ elemental filter; ★ true parameter within the filter bank; ■ true parameter outside of the filter bank.

existing filters — this results in a blended estimate which will be discussed in Section 2.3.4 [198]. The elemental filters should surround³¹ the true parameter [198], as seen in Figure 2.2 when we have a two-dimensional parameter set, i.e., a set of ordered pairs (a_1, a_2) . The open circles represent the location of elemental filters in the bank. The black star is the true parameter at operating point within the filter bank, while the black squares are at points that must be extrapolated since they are not surrounded by the bank of elemental filters.

2. The coarser the discretization, the farther (on average) the true parameter is from the assumed point; in other words, discretization directly affects how

³¹The parameter estimate $\hat{\mathbf{a}} = [\hat{a}_1 \hat{a}_2 \dots \hat{a}_J]^T$ is *surrounded* whenever $\mathbf{a}_\mu \leq \hat{\mathbf{a}} \leq \mathbf{a}_\nu$, i.e., whenever $a_{\mu,j} \leq \hat{a}_j \leq a_{\nu,j}$ for some $\mu, \nu \in \{1, 2, \dots, K\}$ for all J parameters.

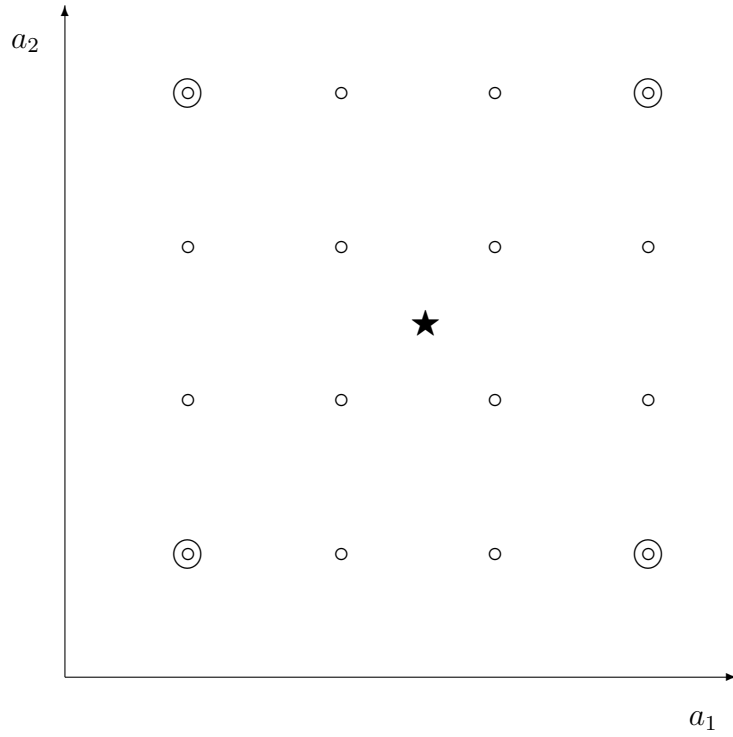


Figure 2.3 Coarse and fine discretization of filter bank. Legend: \bigcirc elemental filters for a coarse discretization; \circ elemental filters for a fine discretization; \star true parameter within the filter bank.

close³² the parameter estimate may be to the true parameter and how well (or how closely) the filter bank surrounds the true parameter [198]. Figure 2.3 shows a coarse discretization of the filter bank using large circles and fine discretization of the filter bank using small circles for a two-dimensional parameter set.

3. The measurement precision (\mathbf{R}) inherently places a lower bound on the practical level of discretization attainable since noisy measurements (a “large” \mathbf{R}) will mask a fine discretization. When multiple filters roughly match the real

³²Closeness can be determined in several ways. Here it refers to “distance” between the true parameter and the assumed parameter used to construct an elemental filter. In a later section, closeness is determined by how accurately the predicted measurement matches the observed measurement; see Section 2.3.4.

world, then the probability flow between them can become unstable when these filters are essentially indistinguishable from one another³³.

4. It is also possible that some intermediate level of discretization can result in a biased estimate since the most appropriate filter may be located “farther” from the true value and the observed parameter value could lie between the assumed parameter values of a pair of filters; however, the estimate is thus biased towards a filter farther from the true parameter value [88].

2.3.3.3 Elemental Filters. Each elemental filter in the bank (shown in Figure 2.1) represents a different system mode and is thus based upon a different hypothesis for the parameter values, e.g., the k th elemental filter design model is constructed assuming that $\mathbf{a}(t_i) = \mathbf{a}_k$. The discrete-time model equations for the k th elemental filter are:

$$\mathbf{x}_k(t_i) = \Phi_k(t_i, t_{i-1}) \mathbf{x}_k(t_{i-1}) + \mathbf{B}_{dk}(t_{i-1}) \mathbf{u}(t_{i-1}) + \mathbf{w}_{dk}(t_{i-1}) \quad (2.38)$$

$$\mathbf{z}(t_i) = \mathbf{H}_k(t_i) \mathbf{x}_k(t_i) + \mathbf{v}_k(t_i) \quad (2.39)$$

where the properties of $\Phi_k(t_i, t_{i-1})$, $\mathbf{B}_{dk}(t_i)$, $\mathbf{H}_k(t_i)$, $\mathbf{Q}_{dk}(t_i)$, and $\mathbf{R}_k(t_i)$ were discussed in Section 2.3.1. Note that most of the research on MMAE has employed the following *time-invariant* system model equations:

$$\mathbf{x}_k(t_i) = \Phi_k(t_i - t_{i-1}) \mathbf{x}_k(t_{i-1}) + \mathbf{B}_{dk} \mathbf{u}(t_{i-1}) + \mathbf{w}_{dk}(t_{i-1}) \quad (2.40)$$

$$\mathbf{z}(t_i) = \mathbf{H}_k \mathbf{x}_k(t_i) + \mathbf{v}_k(t_i) \quad (2.41)$$

where the time-invariance gives $\mathbf{B}_{dk}(t_i) = \mathbf{B}_{dk}$ and $\mathbf{H}_k(t_i) = \mathbf{H}_k$, and for uniform spacing between time samples, $\Delta t = t_i - t_{i-1}$ for all i , the state transition matrix is independent of time: $\Phi_k(t_i, t_{i-1}) = \Phi_k(t_i - t_{i-1}) = \Phi_k(\Delta t)$, and for stationary noises processes: $\mathbf{Q}_{dk}(t_i) = \mathbf{Q}_{dk}$ and $\mathbf{R}_k(t_i) = \mathbf{R}_k$ for all time t_i . These assumptions

³³Note that distinguishability of filters is a function of the filter measurement residuals which is the difference between the predicted and observed measurements as defined in Equation (2.26). Thus we note that parameter set creation by discretization and filter bank composition are closely tied together.

allow the construction of a steady-state Kalman filter model. For the purposes of this research, we will not limit ourselves to steady state filtering.

The appropriateness or validity of each hypothesis, $\mathbf{a}(t_i) = \mathbf{a}_k$, is readily obtained through an analysis of the filter residuals — the difference between the observed measurement and the predicted measurement, $\mathbf{r}_k(t_i) = \mathbf{z}_i - \mathbf{H}(t_i) \hat{\mathbf{x}}(t_i^-)$ [129]. This “correctness” information is encoded in the hypothesis conditional probability $p_k(t_i)$, which is defined as the probability, $\text{pr}\{\cdot\}$, that $\mathbf{a}(t_i)$ assumes the value \mathbf{a}_k (for $k = 1, 2, \dots, K$), conditioned on the observed measurement history to time t_i [130, 132]:

$$p_k(t_i) \triangleq \text{pr}\{\mathbf{a}(t_i) = \mathbf{a}_k | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \quad (2.31)$$

such that

$$p_k(t_i) \geq 0 \text{ for all } k \quad \text{and} \quad \sum_{k=1}^K p_k(t_i) = 1 \quad (2.42)$$

and the mode conditional probability *density* function is actually a conditional probability *mass* function for a discrete random parameter vector [133, 178]:

$$f_{\mathbf{a}(t_i) | \mathbf{Z}(t_i)}(\boldsymbol{\alpha} | \mathbf{Z}_i) = \sum_{k=1}^K p_k(t_i) \delta(\boldsymbol{\alpha} - \mathbf{a}_k) \quad (2.30)$$

If we first assume that the prior probabilities $p_k(t_0)$ are known (or well modeled), for example, $p_k(t_0) = 1/K$ for $k = 1, \dots, K$, then the definition for the hypothesis conditional probability given in Equation (2.31), can be expressed as the following recursion [125, 107, 11, 108, 130, 132, 14]

$$p_k(t_i) = \frac{f_{\mathbf{Z}(t_i) | \mathbf{a}(t_i), \mathbf{Z}(t_{i-1})}(\mathbf{z}_i | \mathbf{a}_k, \mathbf{Z}_{i-1}) p_k(t_{i-1})}{\sum_{j=1}^K f_{\mathbf{Z}(t_i) | \mathbf{a}(t_i), \mathbf{Z}(t_{i-1})}(\mathbf{z}_i | \mathbf{a}_j, \mathbf{Z}_{i-1}) p_j(t_{i-1})} \quad (2.43)$$

where the conditional PDF³⁴:

$$f_{\mathbf{z}(t_i)|\mathbf{a}(t_i),\mathbf{z}(t_{i-1})}(\mathbf{z}_i|\mathbf{a}_k,\mathbf{Z}_{i-1}) = \beta_k(t_i) \exp \left\{ -\frac{1}{2} L_k(t_i) \right\} \quad (2.44)$$

is a zero-mean Gaussian with covariance $\mathbf{A}_k(t_i)$, scale factor

$$\beta_k(t_i) = \frac{1}{(2\pi)^{m/2} |\mathbf{A}_k(t_i)|^{1/2}} \quad (2.45)$$

and measurement dimension m . The likelihood quotient, which is a measure of the “correctness” of the parameter values for this particular model [130], is

$$L_k(t_i) = \mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i) \quad (2.46)$$

where $\mathbf{r}_k(t_i)$ and $\mathbf{A}_k(t_i)$ are the residual and associated residual covariance calculated by the k th Kalman filter as in Equations (2.26) and (2.23), respectively. If we denote the true residual covariance as $\mathbf{A}_{\text{true}}(t_i)$, then $\mathbf{A}_k(t_i) = \mathbf{A}_{\text{true}}(t_i)$ whenever the k th elemental filter properly matches the real world condition. Since the scaling factor, $\beta_k(t_i)$, only ensures that the function always integrates to unity, the important (“shape”) information in this PDF is encoded in likelihood quotient, $L_k(t_i)$ — the weighted square of the residuals.

It has been shown [94, 129] that the sequence of residuals $\{\mathbf{r}_k(t_i)\}$ resulting from linear filtering forms a zero-mean white Gaussian sequence with known residual covariance $\mathbf{A}_k(t_i)$. Thus, if a filter model matches the “true” system, then the residual $\mathbf{r}_k(t_i)$ should be a zero-mean white Gaussian process with known residual covariance $\mathbf{A}_k(t_i)$.

Since we did not derive Equation (2.43), it might not be readily apparent that the denominator is simply the PDF for the current measurement conditioned on the

³⁴Note that conditional PDF $f_{\mathbf{z}(t_i)|\mathbf{a}(t_i),\mathbf{z}(t_{i-1})}(\boldsymbol{\zeta}_i|\mathbf{a}_k,\mathbf{Z}_{i-1})$, where $\boldsymbol{\zeta}_i$ is a dummy variable for the stochastic measurement process, becomes a real number $f_{\mathbf{z}(t_i)|\mathbf{a}(t_i),\mathbf{z}(t_{i-1})}(\mathbf{z}_i|\mathbf{a}_k,\mathbf{Z}_{i-1})$ when evaluated with the measurement $\mathbf{z}(t_i, \omega_j) = \mathbf{z}_i$ at time t_i .

past measurements [210], i.e.,

$$f_{\mathbf{z}(t_i)|\mathbf{z}(t_{i-1})}(\mathbf{z}_i|\mathbf{Z}_{i-1}) = \int_{\mathbb{A}} f_{\mathbf{z}(t_i)|\mathbf{a}(t_i),\mathbf{z}(t_{i-1})}(\zeta_i|\boldsymbol{\alpha},\mathbf{Z}_{i-1}) f_{\mathbf{a}(t_i)|\mathbf{z}(t_{i-1})}(\boldsymbol{\alpha}|\mathbf{Z}_{i-1}) d\boldsymbol{\alpha} \quad (2.47)$$

$$= \sum_{j=1}^K f_{\mathbf{z}(t_i)|\mathbf{a}(t_i),\mathbf{z}(t_{i-1})}(\mathbf{z}_i|\mathbf{a}_j,\mathbf{Z}_{i-1}) p_j(t_{i-1}) \quad (2.48)$$

where the second equality is due to the sifting property of the Dirac delta, as in Equation (2.30). This observation allows us to interpret the discretization of the parameter space into a discrete set of points, each representing a system mode [210]. Recall that the total probability theorem requires that an event, such as the current measurement, be partitioned into a set of disjoint or mutually exclusive partitions such that the union of these partitions equals the event in question [159]. This is in agreement with the discretization guidance given in Section 2.3.3.2. Thus, proper sampling of the parameter space is analogous to proper partitioning of the event space as required by the total probability theorem. In other words, the representation of the parameter space by a discrete set of points is essentially an insightful use of the total probability theorem [210].

2.3.4 State and Parameter Estimates. The MMAE estimation technique uses the information from all of the Kalman filter residuals to estimate the “true” parameter vector in effect and thus determine the true system mode. This technique is optimal when there is a unique filter paired to each of a finite number of system modes. We shall populate the filter bank with K filters; each based on a unique J -dimensional parameter vector.

From the Bayesian point of view, the MMAE framework can be used to compute a state (or parameter) estimate that is characterized by minimizing the MSE between the predicted and measured state estimates; this is most often called a minimum mean-squared error (MMSE) estimate and is the conditional mean. An alternate approach is called the MAP estimate; its estimate corresponds to the largest

hypothesis conditional probability and is dubbed the “closest” model to the real world; the corresponding estimate is the conditional mode. We identify the Bayesian estimate as the standard MMAE estimate and write it as [130, 132, 14]:

$$\hat{\mathbf{x}}_{\text{MMAE}}(t_i^+) \triangleq E\{\mathbf{x}(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i\} = \sum_{k=1}^K \hat{\mathbf{x}}_k(t_i^+) p_k(t_i) \quad (2.49)$$

where $\hat{\mathbf{x}}_k(t_i^+)$ is the state estimate generated by the k th Kalman filter based on the assumption that the parameter vector $\mathbf{a}(t_i) = \mathbf{a}_k$. The conditional covariance of $\mathbf{x}(t_i)$ is [130, 14]

$$\begin{aligned} \mathbf{P}_{\text{MMAE}}(t_i^+) &\triangleq E\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}_{\text{MMAE}}(t_i^+)] [\mathbf{x}(t_i) - \hat{\mathbf{x}}_{\text{MMAE}}(t_i^+)]^T | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \end{aligned} \quad (2.50)$$

$$= \sum_{k=1}^K \{ \mathbf{P}_k(t_i^+) + [\hat{\mathbf{x}}_k(t_i^+) - \hat{\mathbf{x}}_{\text{MMAE}}(t_i^+)] [\hat{\mathbf{x}}_k(t_i^+) - \hat{\mathbf{x}}_{\text{MMAE}}(t_i^+)]^T \} p_k(t_i) \quad (2.51)$$

where $\mathbf{P}_k(t_i^+)$ is the state error covariance computed by the k th Kalman filter. Additionally, the PDF of the state of the system, given the measurement history $\mathbf{Z}(t_i) = \mathbf{Z}_i$, is given by a weighted sum of Gaussian PDFs known as a Gaussian mixture [14]

$$f_{\mathbf{x}(t_i) | \mathbf{Z}(t_i)}(\mathbf{x}_i | \mathbf{Z}_i) = \sum_{k=1}^K \mathfrak{N}[\mathbf{x}(t_i); \hat{\mathbf{x}}_k(t_i^+), \mathbf{P}_k(t_i^+)] p_k(t_i) \quad (2.52)$$

where $\mathfrak{N}[\mathbf{x}(t_i); \hat{\mathbf{x}}_k(t_i^+), \mathbf{P}_k(t_i^+)]$ is the Gaussian (normal) PDF of $\mathbf{x}(t)$ for the k th elemental filter. The parameter estimate is given by:

$$\hat{\mathbf{a}}_{\text{MMAE}}(t_i^+) \triangleq E\{\mathbf{a}(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i\} = \sum_{k=1}^K \mathbf{a}_k p_k(t_i) \quad (2.53)$$

with conditional covariance of $\mathbf{a}(t_i)$ [129]:

$$\begin{aligned} \mathbf{P}_{\mathbf{a}, \text{MMAE}}(t_i^+) &\triangleq E \left\{ [\mathbf{a}(t_i) - \hat{\mathbf{a}}_{\text{MMAE}}(t_i^+)] [\mathbf{a}(t_i) - \hat{\mathbf{a}}_{\text{MMAE}}(t_i^+)]^T | \mathbf{Z}(t_i) = \mathbf{Z}_i \right\} \end{aligned} \quad (2.54)$$

$$= \sum_{k=1}^K [\mathbf{a}_k - \hat{\mathbf{a}}_{\text{MMAE}}(t_i^+)] [\mathbf{a}_k - \hat{\mathbf{a}}_{\text{MMAE}}(t_i^+)]^T p_k(t_i) \quad (2.55)$$

Using the hypothesis conditional probabilities, we can assign a ranking of closeness of the assumed parameter to the true parameter value. Hence the MAP-MMAE state and parameter estimates produced by the Kalman filter with the largest hypothesis conditional probability are given by

$$\hat{\mathbf{x}}_{\text{MAP-MMAE}}(t_i^+) \triangleq \hat{\mathbf{x}}_{k^*}(t_i^+) \quad (2.56)$$

and

$$\hat{\mathbf{a}}_{\text{MAP-MMAE}}(t_i^+) \triangleq \mathbf{a}_{k^*} \quad (2.57)$$

where

$$k^* \triangleq \arg \left\{ \max_k [p_1(t_i), p_2(t_i), \dots, p_K(t_i)] \right\} \quad (2.58)$$

The Bayesian estimates provide smoother transitions as the situation changes, compared to the MAP estimates which may jump from filter to filter. With a properly designed bank of filters, i.e., one which has been properly, if not optimally, discretized, the difference between the Bayesian and MAP estimates is small [68, 139].

2.4 Practical Performance Enhancements for Multiple Model Adaptive Estimation

Depending on the specific problem at hand, there are numerous *ad hoc* adjustments that can be made to the standard MMAE structure to increase either parameter or state estimation performance. Several researchers have compiled lists of some of these techniques. Vasquez [198] compiled his list to improve the composition or positioning of the bank in moving-bank MMAE. Maybeck and Hanlon [135] assembled an assortment of methods to improve sensor/actuator failure detection

and identification. Several other improvement techniques gathered from the literature were not previously considered under the title “enhancement”; however, they will be labeled as such here.

2.4.1 Kalman Filter Tuning. Since no model is exact, the art of filter tuning is the first tool to which we turn after designing a filter. To *tune* an elemental Kalman filter we adjust its process noise covariance \mathbf{Q}_{dk} and/or measurement noise covariance \mathbf{R}_k using an *ad hoc* trial-and-error process [129, 133, 14]. Even though this technique is not analytic, we can use physical insights to help us determine the order of magnitude of the noise covariances. Specifically, one seeks tuning conditions where $L_k(t_i) = \mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)$ is: (1) approximately equal to the number of measurements m when the hypothesized parameter value \mathbf{a}_k is a good match to the true parameter value and (2) significantly larger (or smaller) than m when the model is a poor match. Note that the likelihood quotient is directly affected by the tuning of \mathbf{R}_k since $\mathbf{A}_k = \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k$ and indirectly influenced by \mathbf{Q}_{dk} through the calculation of \mathbf{P}_k^- .

While increasing \mathbf{Q}_d can improve the responsiveness or performance of an individual filter by masking assumed model inadequacies, it may result in slower probability flow between filters because as \mathbf{Q}_d increases, distinguishability among the filters decreases. Too high a value of \mathbf{R}_k can deteriorate the detection capability of the algorithm and will often result in detection delays [130, 132, 47, 46, 71].

2.4.2 Harmonically Balanced Kalman Filters. Muravez [153] designed a harmonically balanced Kalman filter bank to track a target exhibiting a maneuver well modeled by second-order periodically correlated acceleration (PCA) shaping filter in the presence of uncertain measurement noise covariance. The elemental filters were designed to cover the entire frequency range of expected power spectral densities with constant bandwidth filters overlapping at the half-power point. This method as originally proposed requires a large bank of filters and has only been

applied to tracking in a single dimension thus far, i.e., scalar state estimation with scalar measurements. However, it is possible to use a smaller number of filters of the same PCA form with good success [133, 90, 112]. Although this technique is novel, it appears to suffer from having poor distinguishability between the large number of filters required to implement this method “properly” as given in his thesis.

2.4.3 Scalar Residual Monitoring. One way to reduce sensor failure identification ambiguities is through a technique called scalar residual monitoring. The scalar likelihood quotient associated with the j th scalar residual is defined as [139]:

$$L_{k_j}(t_i) \triangleq r_{k_j}^2(t_i) \mathbf{A}_{k_{jj}}^{-1}(t_i) \quad (2.59)$$

which is simply the jj th term of the likelihood quotient³⁵ when written in summation notation:

$$L_k(t_i) = \sum_{\mu=1}^J \sum_{\nu=1}^J r_{k_\mu}(t_i) r_{k_\nu}(t_i) \mathbf{A}_{k_{\mu\nu}}^{-1}(t_i) \quad (2.60)$$

$$= r_{k_j}^2(t_i) \mathbf{A}_{k_{jj}}^{-1}(t_i) + \underbrace{\sum_{\mu=1}^J \sum_{\nu=1}^J}_{\text{except } \mu=\nu=j} r_{k_\mu}(t_i) r_{k_\nu}(t_i) \mathbf{A}_{k_{\mu\nu}}^{-1}(t_i) \quad (2.61)$$

With respect to sensor failure detection and identification, the *predominant indicator* of a failure should be a large value for j th scalar residual $L_{k_j}(t_i)$ in every elemental filter except for the one designed to detect this particular failure³⁶. This makes sense because this term *only* contains information regarding a particular hypothesis about a particular failure [139]. As with many useful *ad hoc* techniques, a threshold must be specified to compare to the scalar likelihood quotient given by

³⁵The likelihood quotient appears in the exponential portion of the hypothesis conditional PDF defined in Equation (2.44)

³⁶Sensor failures are often modeled by zeroing out the row of the output distributor matrix \mathbf{H} corresponding to the failed device and single actuator failures are modeled by zeroing out the appropriate column of the input distributor matrix \mathbf{B} .

Equation (2.59). This technique may be used to augment other methods to identify sensor failures or it may be *the* method to identify them.

2.4.4 β Dominance Compensation. The group led by Athans [10] discovered that the “ β dominance” problem hampered flight condition estimation. Suizu [191] tackled this problem that occurs whenever the elemental filter likelihood quotients are all approximately equal, i.e.,

$$L_1(t_i) \approx L_2(t_i) \approx \cdots \approx L_K(t_i), \quad (2.62)$$

where $L_k(t_i) = \mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)$, then the hypothesis conditional PDF defined in Equation (2.44) is *dominated* by the Gaussian PDF scale factor $\beta_k(t_i) = [(2\pi)^{m/2} |\mathbf{A}_k(t_i)|^{1/2}]^{-1}$; hence the term “ β dominance”. Thus the hypothesis conditional probabilities are inversely related to the determinant of the filter-computed residual covariance $|\mathbf{A}_k(t_i)|$. Hence the model receiving the *largest* probability is the one that has the *smallest* filter-computed residual covariance determinant $|\mathbf{A}_k(t_i)|$. This is totally useless...and worse, it still gives us answers, albeit incorrect ones. For example, a typical representation of a sensor failure is to zero out the row of $\mathbf{H}_k(t_i)$ corresponding to the failed sensor. All other things being equal, the filters designed for this type of failure will tend to have smaller $|\mathbf{A}_k(t_i)|$ values, thus an MMAE devised for sensor failure detection will be prone to false alarms on sensors.

Two simple methods have been implemented to remove this β dominance effect [132, 139, 70, 145, 135, 147, 187, 186]. One method uses scalar residual monitoring and the other simply removes the β_k term from the PDF equation. Thus Equation (2.44) becomes:

$$f_{\mathbf{z}(t_i)|\mathbf{a}(t_i), \mathbf{z}(t_{i-1})}(\mathbf{z}_i|\mathbf{a}_k, \mathbf{Z}_{i-1}) = \exp \left\{ -\frac{1}{2} \mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i) \right\} \quad (2.63)$$

While Equation (2.63) is not strictly a PDF (because the area under the function is not one), the hypothesis conditional probabilities defined in Equation (2.43) still sum to one because of the scaling effect of the denominator.

2.4.5 Scalar Penalty Modification. Another quantity appearing in the definition for the conditional PDF in Equation (2.44) that has been adjusted is the $\frac{1}{2}$ sitting out in front of the likelihood quotient $L_k(t_i)$ [70, 135]. Modifying this number has the same effect as scaling the filter-computed residual covariance inverse \mathbf{A}^{-1} by some scalar “penalty”; with this modification, we are essentially admitting that our PDF is not exactly a Gaussian PDF. By increasing the magnitude, we can raise sensitivity to large residuals, and we can decrease the probability-convergence time for the MMAE. This technique drives the probabilities associated with large residuals to zero faster. Ideally, this should result in a faster convergence of the conditional probabilities; however, increasing the scalar penalty also increases fluctuations in the probabilities and thus results in an increased false alarm rate [70, 135].

2.4.6 Lower Bounding Conditional Probabilities. Placing a lower bound on each of the K hypothesis conditional probabilities, $\{p_k\}$, has been used to prevent filter *lock-out* [10]. While small probabilities result in excessive delays for actuator/sensor failure identification, true filter lock-out *precludes* the identification of the hypotheses associated with the probability that has been set to zero [130, 132, 139, 69, 70, 145, 68, 47, 135, 147, 187, 46, 186]; likewise for other applications that feature abrupt changes in the dynamics of the system such as carrier phase ambiguity resolution [80, 79] and detection of incidents on freeways [212].

Filter lock-out occurs when a hypothesis conditional probability for an elemental filter becomes zero. By inspection of Equation (2.43) one can easily see that once a probability becomes zero at time t_i , then it will remain zero for all time $t > t_i$. This is equivalent to removing the filter from the filter bank altogether. Consequently, it is impossible to identify a failure associated with that particular filter once it has been

“removed.” But with an artificial lower bound in place, a filter cannot be totally locked out and thus it may recover within a “few” iterations and properly declare the failure. The lower bound is an empirically determined value that is greater than zero and for practical purposes, much less than $1/K$. Note that if the lower bound were set at $1/K$, then all of the elemental filters would receive the same probability and hence completely incapacitate the estimator. The larger the lower bound is, the more agile the probability flow is. When a new probability is calculated that would otherwise be less than the lower bound, it is set equal to the lower bound and the entire set of probabilities are re-scaled so that they sum to one.

One drawback to this method is that, when the blended state and parameter estimates are calculated, as shown in Section 2.3.4, they give more weight to the elemental filter estimates which have had their hypothesis conditional probabilities increased by lower bounding. This can be fixed by adding logic to the estimate calculation that simply excludes those filters that have been kept active in the filter bank via lower bounding of computed p_k values, if it becomes problematic. Alternatively, lower bounding does not bias the MAP estimates, provided the lower bound is much smaller than $1/K$ for a bank of K elemental filters.

2.4.7 Markov Process Modeling of Hypothesis Conditional Probabilities.

Ackerson and Fu [2] introduced the concept of modeling the transition from one system mode, represented by parameter vector $\mathbf{a}(t_{i-1})$, to a new system mode, $\mathbf{a}(t_i)$, as a Markov process³⁷. These transitions represent abrupt changes in the dynamics of the system and thus necessitate switching from one elemental filter to another based on a different assumed parameter vector. Under this concept, the hypothesis conditional probabilities are propagated via the Markov process as developed in references [2, 130, 197, 22, 23, 14]. The probability that a Markov system will

³⁷The Markov property allows the conditional PDF for the current parameter value to depend not on the entire time history of parameter values, but on just the previous parameter value, i.e., $f[\mathbf{a}(t_i)|\mathbf{a}(t_{i-1}), \mathbf{a}(t_{i-2}), \dots, \mathbf{a}(t_1)] = f[\mathbf{a}(t_i)|\mathbf{a}(t_{i-1})]$.

transition from mode \mathbf{a}_n to mode \mathbf{a}_m at time t_i , is given by

$$T_{mn}(t_i, t_{i-1}) = \text{pr}\{\mathbf{a}(t_i) = \mathbf{a}_m | \mathbf{a}(t_{i-1}) = \mathbf{a}_n\} \quad (2.64)$$

such that $\sum_{m=1}^K T_{mn}(t_i, t_{i-1}) = 1$. Hence the sequence of modes: $\mathbf{a}(t_0), \mathbf{a}(t_1), \dots, \mathbf{a}(t_i)$ forms a Markov chain. The hypothesis conditional probability vector at time t_i is

$$\mathbf{p}(t_i) = \mathbf{T}(t_i, t_{i-1}) \mathbf{p}(t_{i-1}) \quad (2.65)$$

where the elements, $T_{mn}(t_i, t_{i-1})$, of the $K \times K$ transition probability matrix, $\mathbf{T}(t_i, t_{i-1})$, are given in Equation (2.64) and the hypothesis conditional probability vector

$$\mathbf{p}(t_i) \triangleq \begin{bmatrix} p_1(t_i) \\ p_2(t_i) \\ \vdots \\ p_K(t_i) \end{bmatrix} \quad (2.66)$$

is composed of elements: $p_k(t_i)$ for $k = 1, 2, \dots, K$, previously given in Equation (2.43). The difficult part of utilizing this method is to compute the elements of matrix $\mathbf{T}(t_i, t_{i-1})$ in a meaningful manner; Sullivan and Woodall [192] have proposed a method for “estimating” a Markov state transition matrix. Recently, Jilkov and Li [92] have proposed four algorithms to address the case in which the transition probability matrix is assumed to be time-invariant and random. When \mathbf{T} is the identity matrix, i.e., when the transition probabilities, $T_{mn} = \delta_{mn}$, then the “dynamic” multiple model method becomes a “static” multiple model technique.

The Markov chain assumption obviates the need to employ a lower bounding technique for the hypothesis conditional probabilities, since no matter how small the hypothesis conditional probability, the mode can still “jump” from a high probability to a low probability mode based on the transition probability. Unfortunately, once

we assume that the mode transitions form a Markov chain, the resulting algorithm requires an ever growing amount of memory, hence an optimal algorithm is generally not a feasible option for most applications. However, several suboptimal algorithms have been developed for this Markov-switching concept³⁸: the generalized pseudo Bayesian (GPB) [197, 14] and the interacting multiple model (IMM) [22, 23, 14] are two useful suboptimal methods.

An IMM estimator is an extension of MMAE with Markovian switching that intermixes the state estimates from time t_i to time t_{i+1} in order to approximate the optimal algorithm closely while realizing a huge computational cost savings as compared to the optimal algorithm [22, 23]. Thus, the algorithm trades state estimation accuracy for computational savings.

2.4.8 Hypothesis Swapping. Similar to the hidden Markov modeling of the mode transitions discussed in the previous section, “hypothesis swapping” involves using additional knowledge about how a system operates to help estimate the current mode or parameter. Hoffman [84] used the fact that the T-wave almost always follows a ventricular depolarization and contraction represented by the QRS complex in an electrocardiogram (ECG) signal. Additionally, the T-wave is followed by a variable length “rest” period which is followed by the P-wave. For our purposes here, this knowledge can also be thought of as another form of moving-bank MMAE since only a subset of the entire bank of filters is used at any one time — its composition is modified as necessary as determined by logic rules using information coded in the measurement residuals and other *a priori* information; see Section 2.6.

2.4.9 Probability Smoothing. Immediately following a change in the system operating mode, the probabilities undergo a transition period before converging to the “correct” solution. Probability smoothing is used to minimize the momentary

³⁸See the brief survey paper by Tugnait, reference [197], for two more types of suboptimal algorithms.

false alarms associated with these transients [70, 145, 135, 147, 187, 186]. The probabilities are smoothed over a moving window, i.e., averaged over a number of data samples. The size of the window is chosen empirically: a large window induces a longer delay and a small window allows more false alarms.

2.4.10 Increased Residual Propagation. Another method used to help speed convergence to the best model by skipping a few measurement update cycles while continuing to propagate the Kalman filter state estimates. While still monitoring the filters residuals, they are allowed to grow without the masking affects of measurements. This allows discrepancies between the real world and the model to become more pronounced or visible in the residuals. The number of update samples skipped is determined empirically of course. The risk involved in implementing this technique is an increase in the fluctuations of the conditional probabilities, which gives an increase in false alarms [70, 135]...which can be mitigated by probability smoothing. By skipping measurement updates, we might also degrade the state estimates unless we had an artificially high sampling rate as far as state estimation was concerned.

2.4.11 Dithering. Dithering is the purposeful introduction of periodic (or random) excitation to the system in order to increase the observability of actuator failures through enhanced persistent excitation [145, 146, 58, 57, 147, 187, 71, 186, 73]. Additionally, the sinusoidal dither signal can be explicitly used to identify mismodeled filters. For the properly modeled filter, i.e., the one that matches the real world, the frequency content of the power spectral density (PSD) remains white³⁹, while for the mismodeled filters, a spike appears at the dither frequency [71, 73].

Hanlon [71, 73] harnessed the power of the subliminal dither using a new hypothesis conditional probability calculation to maximize the observability of failed flight control actuators. Since the dither is a highly time-correlated (usually peri-

³⁹Since the dither effect in the observed measurements \mathbf{z}_i is matched by effects in the predicted measurement $\mathbf{H}\hat{\mathbf{x}}(t_i^-)$, the dither is not present in the measurement residuals.

odic) signal, it can be used to mark the filters that are based on a poor model with respect to the real world with a spike at the known dither frequency in the residual’s PSD plot. The residual’s PSD is formed by Fourier transforming the residual’s autocorrelation. Hence, the filter that matches the real world failure will have zero-mean white residuals, while the filters that don’t match the real world failure will pass residual signal power at the dither frequency.

2.4.12 Filter Pruning. For some applications, the parameter space is naturally discrete, hence there does exist an elemental filter that exactly matches the true parameter state. Determining the carrier phase ambiguity for a Global Positioning System (GPS) receiver is one such example [80, 79]. In order to converge on the parameter estimate quickly, we must somehow eliminate or *prune* the filters that are “obviously” incorrect (according to the filter measurement residuals) by incorporating an empirically-based logic rule. Pruning reduces the number of elemental filters in the fixed-bank structure. The distinctiveness of the parameters chosen to represent the parameter set (i.e., the coarseness of the discretization) highly influences what percentage of the filters will be pruned. The filter pruning technique employed by Henderson [80, 79] provides a state estimate that is similar to using the MAP parameter estimate. He cautioned that we must prune carefully so that noisy measurements do not cause the algorithm to delete the “correct” filter mistakenly!

There is a design tradeoff between implementing an exhaustive bank of filters that may have the best answer in it and a small bank of filters that is more computationally feasible. This idea was the impetus for creating a “moving-bank” of filters [130, 81, 136, 132]; the moving-bank MMAE will be developed in Section 2.6. Additionally, the structure of the moving-bank MMAE has been cast in terms of a hierarchical structure [59, 188, 203, 138, 139] discussed in Section 2.7.1 and a filter spawning structure [54, 55], see Section 2.7.2.

For abrupt system mode (or parameter) changes, the number of required hypotheses (or, more correctly, decision tree branches) grows. In Markov-switching parameter systems, the number of hypotheses is K^i at time t_i ! Hence, prudent pruning [4] and/or merging [22, 23] of hypotheses is essential [130].

2.4.13 Filter Restart. When the difference between the predicted measurement $\mathbf{H}_k(t_i) \hat{\mathbf{x}}_k(t_i^-)$ and the observed measurement \mathbf{z}_i increases over time, the filter is said to be *diverging* [130]. Such divergence can occur in practice because, when tuning each elemental filter, one must avoid adding too much pseudonoise to the filter dynamics model (i.e., increase \mathbf{Q}_{dk} too much), since although such conservative tuning can reduce divergence, it can also incapacitate the adaptation in MMAE algorithms. In other words, when the likelihood quotient $L_k(t_i)$ grows without bound and surpasses some threshold, the filter is based on a poor model, i.e., this condition is indicative of a poor match between the real world and the model [105, 106]. As the filter diverges, its output is consequently of little value at best and misleading at worst. Hence, a divergent filter *must* be restarted. One popular method for preventing this situation is to simply re-initialize or restart the divergent filter with the current state estimate, i.e., set $\hat{\mathbf{x}}_k(t_i^-) = \hat{\mathbf{x}}_{\text{MMAE}}(t_i^-)$, where the current state estimate is computed using only the nondivergent elemental filters [133]. Additionally, it may prove useful to restart the covariance estimate $\mathbf{P}(t_i^-)$ as well.

2.4.14 Maximum Entropy with Identity Covariance. Ordinarily, the residuals $\mathbf{r}(t_i^-)$ are weighted by the filter-computed residual covariance $\mathbf{A}(t_i^-)$ when determining the hypothesis conditional PDF $f_{\mathbf{z}(t_i)|\mathbf{a}(t_i), \mathbf{z}(t_{i-1})}(\boldsymbol{\zeta}_i|\mathbf{a}_k, \mathbf{Z}_{i-1})$; see Equations (2.44), (2.45), and (2.46). However, when the filter computed covariance, $\mathbf{A}(t_i^-)$, is suspect, or varies a great deal across the parameter set, \mathbb{A} , then this technique is applied to ensure that the elemental filter with the smallest residual autocorrelation is awarded the highest hypothesis conditional probability, $p_k(t_i)$ as computed in Equation (2.43). We suppress the relative weighting of the residuals by setting the

covariance equal to an $m \times m$ identity matrix [175, 69, 67, 68], i.e.,

$$\mathbf{A}(t_i^-) = \mathbf{I} \quad (2.67)$$

and thus we obtain the maximum entropy with identity covariance hypothesis conditional PDF:

$$f_{\mathbf{z}(t_i)|\mathbf{a}(t_i),\mathbf{z}(t_{i-1})}(\mathbf{z}_i|\mathbf{a}_k,\mathbf{Z}_{i-1}) = \frac{1}{(2\pi)^{m/2}} \exp \left\{ -\frac{1}{2} \mathbf{r}_k^T(t_i) \mathbf{r}_k(t_i) \right\} \quad (2.68)$$

Sheldon called this modification to the algorithm maximum entropy with identity covariance (ME/I) since it maximizes the entropy of the residual information [175].

2.4.15 Pseudo-Residuals. A novel method for detecting measurement bias jumps (such as GPS spoofing [206, 205, 204]) uses a pseudo-residual vice the true measurement residual as defined in Equation (2.26). The pseudo-residuals are used only for inspecting the residuals while the true residuals are used to update the elemental filters. While true residual sequences (at steady state) are zero-mean white Gaussian sequences with known covariances, the pseudo-residual sequences have a nonzero mean equal to the assumed bias. Hence this formulation allows the bias to be detected.

2.4.16 Generalized Residuals. For some applications, one might conjecture that the MMAE may benefit from a different form for the measurement residual. For instance, when the uncertain parameter (\mathbf{a}) affects the measurement model's structure (\mathbf{H}) and/or statistics (\mathbf{R}) and propagation errors dominate the state estimate [80, 79], then perhaps an analysis of the “post-fit” residuals

$$\mathbf{r}(t_i^+) \triangleq \mathbf{z}_i - \mathbf{H}(t_i) \hat{\mathbf{x}}(t_i^+) \quad (2.69)$$

with an error covariance of:

$$\mathbf{A}(t_i^+) \triangleq \mathbf{H}(t_i) \mathbf{P}(t_i^+) \mathbf{H}^T(t_i) + \mathbf{R}(t_i) \quad (2.70)$$

may be the best method for adaptively estimating the parameter — compare to the standard forms given in Equations (2.26) and (2.23), respectively. That is, the distinguishability of the elemental filters is (assumed to be) more evident through an analysis of the post-fit residuals than with the standard set. In a more detailed analysis, Ormsby [156] showed that these post-fit residuals actually resulted in no performance improvement when compared to an MMAE using traditional residuals. In showing this he constructed a generalized residual as a weighted sum of the traditional and post-fit residuals:

$$\mathbf{r}^*(t_i) = \gamma \mathbf{r}(t_i^-) + (1 - \gamma) \mathbf{r}(t_i^+), \quad (2.71)$$

where the scalar γ is chosen by the designer to optimize the performance. When $\gamma = 1$ we have the traditional residual and when $\gamma = 0$ we have the post-fit residual. There is no theory regarding how to determine an optimal γ , however, it is suspected that the optimal γ is a number between zero and one.

Ormsby [156, 157] showed that previous researchers [80, 79] who used the post-fit residuals would have gotten equivalent results using the traditional form of the residuals. One side effect of the generalized residual, for $\gamma \neq 1$, is the beta dominance effect which was previously discussed in Section 2.4.4; hence researchers must be careful when choosing the weighting factor.

2.5 A Virtual Filter Bank Using Only a Single Filter

Hanlon and Maybeck [71, 72] have proposed a computation-saving virtual filter bank using a *single* Kalman filter combined with a set of linear transformations,

rather than a set of K elemental filters⁴⁰. The linear transformations capture the differences between the model used for the single filter and the models for the virtual filters. The virtual filter bank is composed of a bank of linear transformations that compute equivalent state estimates and residuals based on a single Kalman filter which produces the reference state estimate and measurement residual. They have developed the necessary linear transforms to model differences in the input distributor matrix \mathbf{B}_d ⁴¹, output or measurement distributor matrix \mathbf{H} , and the state transition matrix Φ . Their development is based on the time invariant model as in Equations (2.16) and (2.17). Additionally, they have assumed that the Kalman filter models and the truth model dynamics noise covariance \mathbf{Q}_d , measurement noise covariance \mathbf{R} , and noise distributor matrices \mathbf{G}_d are equivalent. They indicate that these conditions are common in failure detection applications for which the MMAE methodology has been used.

The development begins by rewriting the measurement residuals as defined in Equation (2.26) for the j th filter with the goal of eliminating the explicit mention of the state estimate from the j th filter. After several lines of algebra, they write:

$$\begin{aligned} \mathbf{r}_j(t_i) = & \mathbf{r}_k(t_i) + \mathbf{H}_j \Phi_j \Delta \epsilon_{jk}(t_{i-1}^+) + [\mathbf{H}_j \Delta \Phi_{kj} + \Delta \mathbf{H}_{kj} \Phi_k] \hat{\mathbf{x}}_k(t_{i-1}^+) \\ & + [\mathbf{H}_j \Delta \mathbf{B}_{kj} + \Delta \mathbf{H}_{kj} \mathbf{B}_k] \mathbf{u}(t_{i-1}) \end{aligned} \quad (2.72)$$

where the difference in the state estimation errors for the j th and k th filters, $\epsilon_{jk}(t_i^+)$, is algebraically equivalent to the difference in the state estimates

$$\Delta \epsilon_{jk}(t_i^+) \triangleq \epsilon_j(t_i^+) - \epsilon_k(t_i^+) = \hat{\mathbf{x}}_k(t_i^+) - \hat{\mathbf{x}}_j(t_i^+) \quad (2.73)$$

⁴⁰While this section does not directly support the work documented in this dissertation, it is an example of potentially useful insights.

⁴¹For the purposes of this section, the subscript d — which is simply a reminder of the discrete nature of the quantity — will be suppressed in the discussion following this introductory paragraph.

which can be expressed using the following recursion:

$$\begin{aligned}
\Delta\epsilon_{jk}(t_i^+) &= (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)\Phi_j\Delta\epsilon_{jk}(t_{i-1}^+) \\
&+ [(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\Phi_k - (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)\Phi_j]\hat{\mathbf{x}}_k(t_{i-1}^+) \\
&+ [(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{B}_k - (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)\mathbf{B}_j]\mathbf{u}(t_{i-1}) + \Delta\mathbf{K}_{kj}\mathbf{z}(t_i)
\end{aligned} \tag{2.74}$$

and finally, four equivalence relations

$$\begin{aligned}
\Delta\mathbf{B}_{kj} &\equiv \mathbf{B}_k - \mathbf{B}_j \\
\Delta\mathbf{H}_{kj} &\equiv \mathbf{H}_k - \mathbf{H}_j \\
\Delta\Phi_{kj} &\equiv \Phi_k - \Phi_j \\
\Delta\mathbf{K}_{kj} &\equiv \mathbf{K}_k - \mathbf{K}_j
\end{aligned} \tag{2.75}$$

While the linear transforms represented by Equations (2.72) and (2.74) appear to be very complex, in practice, they often simplify drastically — at least for the failure detection application that Hanlon and Maybeck have addressed.

2.5.1 An Example: Different Input Distributor Matrices. To create an actual filter to model a single actuator failure, we would zero out a column of the input distributor matrix \mathbf{B} . If we assume that the only difference between the reference model (call it the k th filter) and this one used to detect an actuator failure in the j th filter, then of the four increment matrices given in Equation (2.75) only $\Delta\mathbf{B}_{kj}$ is nonzero, hence Equations (2.74) and (2.72) simplify to

$$\Delta\epsilon_{jk}(t_i^+) = (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)\Phi_j\Delta\epsilon_{jk}(t_{i-1}^+) + (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)\Delta\mathbf{B}_{kj}\mathbf{u}(t_{i-1}) \tag{2.76}$$

and

$$\mathbf{r}_j(t_i) = \mathbf{r}_k(t_i) + \mathbf{H}_j\Phi_j\Delta\epsilon_{jk}(t_{i-1}^+) + \mathbf{H}_j\Delta\mathbf{B}_{kj}\mathbf{u}(t_{i-1}) \tag{2.77}$$

Hanlon and Maybeck [71, 72] have developed similar relations for the case of a single sensor failure (different measurement distributor matrices, i.e., $\Delta \mathbf{H}_{kj} \neq \mathbf{0}$, which also means that the Kalman gains are likely to be different, hence $\Delta \mathbf{K}_{kj} \neq \mathbf{0}$) and for different state transition matrices $\Phi_k \neq \Phi_j$ (again $\Delta \mathbf{K}_{kj} \neq \mathbf{0}$).

2.5.2 Equivalent Residuals. While we won't reproduce their work here, we must note that they spent considerable effort showing that the difference in the equivalent residual produced by Equation (2.72) is essentially identical to the measurement residual from the actual filter — nominally within the precision of the simulation software. Thus, this technique is viable provided that the extra work necessary to set up the new algorithmic apparatus is more than offset by the reduced computational load provided by this framework.

2.5.3 Computational Savings. While results will vary depending on the particular application, they have estimated (using an operations counting technique) that this virtual filter bank design can yield savings of about 30% for the case of different input distributor matrices as in the example above. Similarly, they found that the equivalent residual version of the Kalman filter bank reduces the required operations count by about 15% compared to the fully implemented Kalman filter bank. It is noted that there are basically no savings if the differences are confined to the state transition matrices.

2.5.4 Comment. While the thrust of the current research is to extend the MMAE, one can not overemphasize the practical matter of reducing the computational load whenever possible. Aside from the utility of this formulation, Hanlon and Maybeck have noted the similarity of this work to the generalized likelihood ratio test that also uses the residuals from a single Kalman filter to detect failures [211, 71, 72, 198].

2.6 Moving Bank Multiple Model Adaptive Estimation Fundamentals

Until now, the system mode (parameter vector) was assumed to be time-invariant for static fixed bank MMAEs. If the system mode is allowed to vary slowly with time over a large set of admissible parameter vectors, then a finely discretized parameter space may give rise to a prohibitively large collection of elemental filters. Since the number of filters kept on-line is generally limited, we are motivated to consider methods capable of adjusting the composition of the filter bank. Maybeck [130] suggested an *ad hoc* approach to track slowly varying parameter vectors via a “dynamically redefinable (or ‘moving’) bank of filters” as opposed to a fixed “static” bank of filters. With this technique, the bank of filters can be both closely spaced (a necessary condition for producing good state/parameter estimates) and still cover the entire range of system modes while adhering to the constraint of keeping only K elemental filters online at any point in time.

Hentz and Maybeck [81, 136] completed the first feasibility study of this “moving bank” MMAE structure. The moving-bank MMAE estimates the parameter vector using only a small subset of the entire filter bank. Subsequent investigations were undertaken by Maybeck, Gustafson, Griffin, Schiller, Vasquez, and Erickson [132, 172, 68, 64, 198, 200, 199, 49, 201, 50] to reduce the number of on-line filters required in the MMAE bank. Li and Bar-Shalom [120] have proposed a similar architecture for the IMM called variable structure; this enhancement allows the model-set to be dynamically redeclared online.

If the parameter estimate changes appreciably, as indicated by the filter residual statistics, then the bank of on-line filters should be adjusted so that the parameter estimate is always surrounded by a “moving” bank of elemental filters⁴². Hence a primary purpose for changing the elemental filters used in the MMAE bank is to track the true parameter vector value using a small number of filters with fine dis-

⁴²The subset of filters appears to *move* through the parameter set as the parameter changes while time unfolds.

cretization of the parameter space to reduce the number of filters that must be kept on-line. A second reason for changing the elemental filters used in the MMAE filter bank is to acquire (or reacquire) the true state vector using a coarse discretization of the entire parameter space; this may occur multiple times during the course of a simulation or actual use. The composition of this moving bank is governed by a set of logic rules that will be discussed shortly.

2.6.1 *Moving-Bank versus Fixed-Bank Multiple Model Adaptive Estimation.*

The moving-bank MMAE posed by Maybeck and Hentz is the same as the filter bank estimator discussed previously except that K now refers to a smaller number of elemental filters in the moving bank rather than the total number of possible elemental filters based on all possible discrete parameter vector values, usually $K \ll K_{\text{fixed-bank}}$, where $K_{\text{fixed-bank}}$ is the total number of filters in the reservoir known as the fixed bank. Usually the bank of K elemental filters “moves” within a previously constructed fixed bank of elemental filters such that each $\mathbf{a}_k \in \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{K_{\text{fixed-bank}}}\} \subset \mathbb{A}$, for $k = 1, \dots, K$, as was originally implemented and investigated [132, 172, 68, 64]; also, the K elemental filters may be created on-line and may “roam” throughout the entire parameter space, i.e., $\mathbf{a}_k \in \mathbb{A}$ by discretizing the parameter set on-line as reported by Miller, Vasquez, and Maybeck [149, 198, 200, 199, 201]. Several decision logics designed to reassign the K “mobile” filters from the larger set of $K_{\text{fixed-bank}}$ filters have been suggested and are presented later in this section. Several of these rules are also used by the dynamic bank implementation created on-line based on a modified formulation of Sheldon’s optimal parameter discretization strategy [175, 176, 177].

In addition to moving the filter bank while we track the parameter changes, we may also expand the filter bank region of coverage⁴³ when it appears, i.e., when

⁴³When the region of coverage is expanded, the number of elemental filters in the filter bank is usually held constant. The IMM variable structure approach [120] is one method that allows the number of elemental filters to vary.

the statistics of the filter residuals suggest it, that the true parameter value lies outside the bounds of the current bank. To expand the bank, we simply increase the coarseness of the discretization. This will hopefully place the true parameter value within the confines of the moving bank once again; i.e., we'd like to surround the parameter we are trying to estimate. Note that moving the bank and expanding the bank are two separate decisions.

Once a bank has been expanded in order to recapture the true value, the quality (or accuracy) of the parameter estimate falls off (assuming that the number of filters in the bank is kept constant) until we contract the bank. In other words, it is important that the true parameter lie within the bank of filter's coverage area for adequate parameter estimation. This will increase our chances that one of our elemental filters is "close" to the true parameter value. The level of discretization of the continuous parameter set directly impacts the ability of the filter bank to surround and come as close as possible to the true parameter value.

Many researchers have proposed novel ways of adjusting the bank of elemental filters, e.g., the new set of rules developed by Vasquez and Maybeck [198, 199, 200, 201]. Additionally, several authors have proposed different "moving-bank" architectures for modifying the bank of filters. For example, the "filter spawning" architecture [54, 55] will be discussed in Section 2.7.2, while the variable structure approach designed by Li and Bar-Shalom [120, 118] is outside the scope of this research.

2.6.2 A Short Glossary of Bank Manipulation Terms. Before proceeding with a review of the logic rules used to control the composition of the filter bank, we shall discuss a few important terms regarding the movement of the moving-bank MMAE framework:

Contract The filter bank *contracts* when the discretization level is made finer (assuming that the same number of filters is maintained in the bank.)

Expand The filter bank *expands* when the discretization level is made coarser (assuming that the same number of filters is maintained in the bank.)

Move To *move* the filter bank is to re-center the bank on the latest parameter estimate while the discretization level is (possibly) held constant.

Surround The filter bank *surrounds* the true parameter value when the bank contains filters that bound the parameter estimate both above and below in all “directions” of the parameter set.

Track The filter bank *tracks* the true parameter value by keeping the parameter estimate surrounded at all times; this may require the move, expand, and contract filter bank operations.

2.6.3 Logic Rules for Moving the Bank. Five standard decision logics have been suggested and investigated by Maybeck [132] and others in order to keep the estimate of the parameter within the bounds of the bank. A brief summary of each follows. Additionally, Vasquez and Maybeck [198, 199, 201] have developed an algorithm that exploits the information contained in a conditional PDF to move the filter bank.

2.6.3.1 Residual Monitoring. The likelihood quotient defined in Equation (2.46):

$$L_k(t_i) = \mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i), \quad (2.46)$$

captures the useful information pertaining to the correctness of the parameter values. For scalar measurements, this is simply the current residual squared, divided by the filter computed variance for the residual: $L_k(t_i) = r_k^2(t_i)/A_k(t_i)$. When the true parameter value does not lie in the current moving-bank region, all K likelihood quotients can be expected to exceed a threshold level T_L , the numerical value of

which is set in an *ad hoc* manner during performance evaluations⁴⁴. Thus, a possible detection logic would indicate that the bank should be either moved or expanded at time t_i if

$$\min\{L_1(t_i), \dots, L_K(t_i)\} \geq T_L \quad (2.78)$$

In other words, we should expand the moving bank when all of the likelihood quotients are too large. If we apply the test in Equation (2.78) to only a specific subset of the current bank, we may infer movement of the true parameter and thus direct that the bank be moved in order to track the true parameter. For example, if all of the likelihood quotients for the filters along the bank's edge exceed some threshold, then the appropriate action would be to move the bank in the opposite direction where the likelihood quotients are smaller. This tool is prone to false alarms because it is based on a single residual at time t_i . We could lower false alarms by averaging over several time samples, but this strategy would tend to decrease our responsiveness to real world changes in the parameter, which, as you will recall, is the hallmark of this method. Hence we need a better logic rule.

2.6.3.2 Parameter Position Estimate Monitoring. Since our intent may well be to track the actual value of the parameter through the parameter set, we could explicitly monitor the actual position estimate of the parameter as a function of time. Recall that the Bayesian MMAE parameter estimate given by Equation (2.53) is⁴⁵

$$\hat{\mathbf{a}}(t_i) = E\{\mathbf{a}(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i\} = \sum_{k=1}^K \mathbf{a}_k p_k(t_i) \quad (2.79)$$

If the difference between the parameter estimate $\hat{\mathbf{a}}(t_i)$ and the “center” of the filter bank, $\mathbf{a}_{\text{center}}$, becomes too large, i.e., larger than some chosen threshold, then the filter should be moved in such a manner as to bring the center $\mathbf{a}_{\text{center}}$ and $\hat{\mathbf{a}}$ closer

⁴⁴See Hanlon's [71] work on Neyman-Pearson hypothesis testing process as an alternative for the ad hoc method with which the threshold is set.

⁴⁵We have suppressed the difference between the propagated and update estimate here since this is not germane to the discussion.

together. Since $\hat{\mathbf{a}}(t_i)$ depends on a *history* of measurements rather than just the single current measurement, it is less prone to the false alarms compared to the simple residual monitoring method discussed above.

2.6.3.3 Parameter Position and Velocity Estimate Monitoring.

Another way to incorporate more data into our decision making logic is to use the history of $\hat{\mathbf{a}}(t_i)$ to generate a meaningful estimate of the parameter velocity⁴⁶ if the true parameters vary slowly with time — otherwise we will get what may appear to be as random motion:

$$\dot{\hat{\mathbf{a}}}(t_i) \cong \frac{\hat{\mathbf{a}}(t_i) - \hat{\mathbf{a}}(t_{i-1})}{t_i - t_{i-1}} \quad (2.80)$$

The parameter estimate velocity $\dot{\hat{\mathbf{a}}}(t_i)$ and current position estimate $\hat{\mathbf{a}}(t_i)$ can be used to predict the parameter position one sample period into the future by rearranging Equation (2.80):

$$\hat{\mathbf{a}}(t_{i+1}) \cong \hat{\mathbf{a}}(t_i) + \dot{\hat{\mathbf{a}}}(t_i)[t_{i+1} - t_i] \quad (2.81)$$

If the distance between the bank center and that prediction $\hat{\mathbf{a}}(t_{i+1})$ exceeds some selected threshold, then the bank can be moved in anticipation of the true parameter movement. This approach introduces lead into the moving-bank logic, but also a higher level of uncertainty and possibly erratic bank movement if the true value of the parameter changes too rapidly.

2.6.3.4 Probability Monitoring.

The conditional hypothesis probabilities, $p_k(t_i)$, computed using Equation (2.43), are another indication of the correctness of the parameter values \mathbf{a}_k assumed by the elemental filters of the current bank. If any of these probabilities rise above a chosen threshold level, the bank can be moved in the direction of the \mathbf{a}_k associated with the highest $p_k(t_i)$. In this scheme, the bank seeks to center itself on the elemental filter with the highest conditional probability weighting. Again, since $p_k(t_i)$ depends on a history of measurements,

⁴⁶Velocity is simply the time rate of change of position.

this method should be somewhat insensitive to singular instances of measurement corruption as is the case under residual monitoring.

2.6.3.5 Parameter Estimation Error Covariance Monitoring.

Whereas residual monitoring may be used to increase the spacing of the filters in the filter bank, i.e., expand the bank, this technique allows us to *contract* the bank by decreasing the discretization level of the parameter set. By starting with a coarsely discretized parameter set, $\mathbf{a}_1, \dots, \mathbf{a}_K$, we increase our chances of surrounding the true parameter value. Then with proper testing we may contract the bank centered on that parameter estimate. A good way to help make such a contraction decision is to monitor the parameter estimation error conditional covariance [130] given in Equations (2.54) and (2.55) and repeated here (minus the “MMAE” subscripts)

$$\mathbf{P}_{\mathbf{a}}(t_i^+) \triangleq E \{ [\mathbf{a}(t_i) - \hat{\mathbf{a}}(t_i^+)] [\mathbf{a}(t_i) - \hat{\mathbf{a}}(t_i^+)]^T | \mathbf{Z}(t_i) = \mathbf{Z}_i \} \quad (2.82)$$

$$= \sum_{k=1}^K [\mathbf{a}_k - \hat{\mathbf{a}}(t_i^+)] [\mathbf{a}_k - \hat{\mathbf{a}}(t_i^+)]^T p_k(t_i) \quad (2.83)$$

When an appropriately chosen norm (a scalar function indicating size or distance) of this matrix falls below a selected threshold, the bank can be contracted about the parameter estimate. One such norm is the weighted sum of the diagonal terms of the covariance matrix $\mathbf{P}_{\mathbf{a}}(t_i)$ for a moving bank constrained to be a square region in a two-dimensional parameter set [132]:

$$||\mathbf{P}_{\mathbf{a}}(t_i)|| \triangleq \sum_{k=1}^K [\mathbf{P}_{\mathbf{a}}]_{kk}(t_i) \quad (2.84)$$

In general, one could use different discretization coarseness decisions in individual directions of the parameter set, allowing rectangular banks as well as squares.

Note that we can rewrite the state equations for a deterministic time-invariant system⁴⁷ into diagonal *canonical* form if the eigenvalues of the system transfer function are unique (non-repeated) or into a modified canonical form if there are repeated roots. Similarly, we can use the eigenvalues and eigenvectors of the covariance matrix $\mathbf{P}_{\mathbf{a}}(t_i)$ at time t_i to find the principal axes. With these principal axes, we could describe elliptical shaped banks in the parameter set [129].

An indication of the need to expand the size of the bank can be obtained from residual monitoring as before. When all of the likelihood quotients from Equation (2.46) are large in magnitude (indicating that none of the current elemental filters appear to have a good model or hypothesized parameter value), then it is more appropriate to expand the bank than to attempt to move it because no clear indication of the true parameter's value is provided with this particular bank of filters. The error covariance could then be monitored for making the decision to return the bank to a smaller size. Since Equation (2.83) depends on the current choice of \mathbf{a}_k values, this error covariance is not a reliable indicator for the decision to expand because the computed $\mathbf{P}_{\mathbf{a}}(t_i)$ is artificially bounded above by the current size of the bank. Regardless of which technique is used to move, contract, or expand the filter bank, the newly declared models \mathbf{a}_k must be initialized with values for $\hat{\mathbf{x}}_k(t_i)$, $\mathbf{P}_k(t_i)$, and $p_k(t_i)$. A common and reasonable choice for $\hat{\mathbf{x}}_k(t_i)$ is the current moving-bank blended estimate $\hat{\mathbf{x}}(t_i^+)$. For the new $p_k(t_i)$, we equally divide up the probability weight of the discontinued filters, i.e., if filters one through three had a total probability of just one tenth, then each of the new filters will have probability of one

⁴⁷The deterministic time-invariant system is

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{F}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{z}(t_i) &= \mathbf{H}\mathbf{x}(t)\end{aligned}$$

where \mathbf{F} , \mathbf{B} , and \mathbf{H} are as described in Section 2.3.1. The system transfer function matrix for this system is

$$\mathbf{G}(s) = \mathbf{H}^T[s\mathbf{I} - \mathbf{F}]^{-1}\mathbf{B} \quad (2.85)$$

where s is the Laplace transform variable [129].

third of one tenth. Another method apportions the probability based on the relative correctness of the new filters being added [81, 136]. This correctness is based on an evaluation of the likelihood quotients determined from Equation (2.46) for each new filter; it is used to divide the probability proportionately. This method equates the smallest likelihood quotient with the most correct filter and thus the most correct filter shall receive the greatest probability allocated to the new filters. However, this apportionment technique doesn't usually perform better in practice relative to simpler equal-distribution methods.

2.6.3.6 *Density Algorithm: Logic Rules for Moving the Bank.* As

opposed to the simple rules just discussed in the preceding subsections, the “density algorithm” developed by Vasquez and Maybeck [198, 199, 201] provides intelligent decision making for movement, contraction, and expansion of the adaptive MMAE filter bank. The density algorithm gets its name by exploiting information provided by the hypothesis conditional probability *density* function $f_{\mathbf{z}(t_i)|\mathbf{a}(t_i), \mathbf{z}(t_{i-1})}(\boldsymbol{\zeta}_i|\mathbf{a}_k, \mathbf{Z}_{i-1})$ defined in Equation (2.44). Unfortunately, this algorithm relies heavily on uniform spacing of the parameter values of the online filters. Note that uniform parameter value spacing is not a usual feature of the bank composition. This prompted Vasquez [198] to combine the basic density algorithm with a new online discretization technique that does not rely on simple uniform spacing of the parameter values.

2.6.4 *Hypothesis Testing.* Multiple model estimation employs hypothesis testing in a variety of ways, using the filter measurement residuals to help make decisions. Hypothesis testing of the residuals (or more commonly, a function of the residuals) is used to determine when the composition of the bank should be changed and how it should be modified. The testing of the hypotheses is how changes in the system are detected; this is known as detection theory, see, e.g., [170, 101]. The multiple model estimation schemes discussed in this document would not be possible without hypothesis testing.

The proximity of the parameter estimate to the true value is dependent on the coarseness of the discretization. Larger than expected residual magnitudes indicates a mismatch between the filter model and the “truth” model. The truth model is the best model that we can build irrespective of its feasibility of employment — we desire the closest possible representation of the real world system. A parameter change in the true system would be reflected as a change in magnitude of the residuals of filters based on different hypotheses when the change occurred. The change in the residuals can appear as a nonzero mean or a change of covariance.

In addition to the hypothesis conditional probability computed for each model, see Equation (2.43), we can also gather additional information about the models through decision theory via hypothesis testing. Hypothesis testing is a process of establishing the validity of a hypothesis, where the statistical hypothesis is simply an “assumption about the value of one or more parameters of a statistical model” [159] or more precisely, an “assertion about the [probability] distribution of one or more random variables” [86]. The null hypothesis is usually the condition of no change while the alternative hypothesis encapsulates the “change.” The hypothesis test is a rule which can be used to determine whether or not to reject the null hypothesis⁴⁸.

The two common classes of errors are the type I and type II errors. Type I errors occur when we mistakenly reject a null hypothesis that is actually in force (true); a type II error happens when we fail to reject (or accept) a null hypothesis that is not in force, i.e., false [168]. For our work, the null hypothesis occurs when the system is functioning as intended, while a malfunctioning system would be the alternate hypothesis; as the name suggests, there may be more than one alternate hypothesis in a multiple hypothesis testing scheme. When a hypothesis test falsely indicates that the system is malfunctioning, a type I error has occurred — this error is commonly known as a false alarm. Similarly, a type II error is committed whenever

⁴⁸Note that if a hypothesis is *not rejected* that it is not necessarily true, this *acceptance* merely indicates that the data considered supports the conjecture; likewise, a *rejected* hypothesis is not necessarily false [168].

the test indicates that the system is functioning properly, when in fact it is operating “poorly” in some respect; this error is often called a missed detection in that the test missed detecting the problem. When applied to the target detection problem, a false alarm simply says that there was a target present when none truly was, and a missed detection is simply that the target was present, but that the test indicated a target free area.

2.6.4.1 Chi-Squared Test. The chi-squared test is one of the earliest methods used for statistical inference [86]; in failure detection, it uses the measurement residuals from a (Kalman) filter to determine whether a failure has occurred. Thus, this test gathers knowledge of the system dynamics by interpreting the filter measurement residuals. The chi-squared random variable, $\chi^2(t_i)$, provides a test statistic that places a quadratic penalty on residual variance for the k th Kalman filter model [198]:

$$\chi^2(t_i) = \sum_{l=i-N+1}^i \mathbf{r}_k^T(t_l) \mathbf{A}_k^{-1}(t_l) \mathbf{r}_k(t_l) \quad (2.86)$$

where N is the size of the sliding window used to make the decision. A detection rule with an empirically determined threshold T_{χ^2} is:

$$\begin{aligned} \chi^2(t_i) &> T_{\chi^2} \rightarrow \text{Parameter Change} \\ \chi^2(t_i) &\leq T_{\chi^2} \rightarrow \text{No Parameter Change} \end{aligned} \quad (2.87)$$

The threshold will be chosen to meet the performance specifications and to minimize false alarms (type I errors) and missed detections (type II errors) [170]. The chi-squared test has been a highly effective and consistent failure detector [152]; however, this test is basically an alarm method that isn’t of much use for isolating failures if not used in a multiple model structure [211].

2.6.4.2 Neyman-Pearson Test. The most powerful test is defined as the hypothesis test which yields the greatest probability of detection for a given

level of false alarms. The Neyman-Pearson (hypothesis) test, based on the Neyman-Pearson lemma, yields the most powerful test [170]. Hanlon [71] replaced the standard N -ary hypothesis test with a Neyman-Pearson based hypothesis test extended especially for the MMAE structure. In an $N = 2$ binary hypothesis test, the null hypothesis is either rejected or accepted, whereas in a Neyman-Pearson test, a third option located “between” acceptance and rejection of the null hypothesis is available. The somewhat noncommittal third response is to reject the null hypothesis with a certain probability as determined by the test function; thus acceptance of the null hypothesis is the same as rejection with zero probability. With Hanlon’s Neyman-Pearson hypothesis testing algorithm [71], the residual sequence from a single Kalman filter could be used to perform multiple hypothesis tests, as opposed to having an entire bank of filters feeding an N -ary hypothesis testing algorithm used by the standard MMAE.

2.6.4.3 Sequential Probability Ratio Test. For “short” fixed-length data sets, there may not be enough information to distinguish between the various hypothesis. In this case, it is desirable to have a test that continues to collect data until there is enough information to make a decision. When this occurs, the sequential probability ratio test (SPRT) developed by Wald [202] is superior to the Neyman-Pearson test [120].

2.6.4.4 Generalized Likelihood Ratio Test. The generalized likelihood ratio (GLR) test is similar to the chi-squared test, but it has the capability to detect abrupt changes in dynamic systems and isolate failures [214, 211, 213, 212, 152, 206, 198, 101]. A hypothesis for each type of failure that can affect the system is constructed. The GLR test processes the measurement residuals from a single Kalman filter in parallel in order to detect changes (failures) in the system, whereas in MMAE, we have a bank of Kalman filters. One key benefit of the GLR is that it needs only one estimator for each failure type since it produces its own estimate of the

failure magnitude. This magnitude may be used in a feedback loop to aid the system in terms of recovery or reconfiguration. The MMAE has generally outperformed the GLR test.

2.7 *Advanced Moving-Bank Multiple Model Adaptive Estimation Structures and Techniques*

In this section, we shall briefly discuss several advanced/specialized techniques beginning with two moving-bank-like structures that have been applied to the failure detection problem. Next we'll cover two algorithms for discretizing the parameter space. Then we'll investigate a modification to the hypothesis conditional PDF which results in a hypothesis testing algorithm designed not only to detect failures, but to isolate them. Finally, we briefly introduce a method for improving state estimation by using an MMAE to identify the system mode, i.e., the unknown system parameter, and then using a separate Kalman filter for state estimation.

2.7.1 Hierarchical Structured Filter Bank. Hierarchical estimation theory was studied by Smith and Sage [179], blended with Magill's multiple model method [125] by Fry and Sage [59], and later implemented by Stevens, Maybeck, and others [188, 138, 139, 145, 147, 47, 46, 35] as a way to reduce the required number of elemental filters to detect multiple actuator/sensor failures for reconfigurable flight control via MMAE methods. If a multiple model algorithm were based on all possible single and double failures of N sensors and actuators, then we would need filters to correspond to the cases of: Fully functional (1 filter), single failure (N filters), and double failures ($\frac{N!}{(N-2)!2!}$ filters⁴⁹) for a total of $1 + N + \frac{N!}{(N-2)!2!}$ elemental filters. To avoid this computational burden, we could cast the problem into a hierarchical structure where we only have $N + 1$ on-line filters at any given time. The initial bank of filters are denoted by the title "Level Zero" which corresponds to zero failures

⁴⁹Recall that for positive integer N , the factorial is the product: $N! = N \cdot (N - 1) \cdots 1$.

detected, i.e., the models are based on the condition that no failures have been detected. We also have N banks of $N + 1$ filters on standby; the “Level One” banks. Upon confirmation of an initial failure, we bring on-line the appropriate bank designed with the assumption that a failure has occurred in the n th sensor or actuator surface. In this “Level One” bank, we also include the “no failure” filter to handle the possibility that the sensor (or actuator) was mistakenly declared inoperative or allow for the possibility that it can get better; hence $N + 1$ filters. So, at most, we will have 11 filters online in the case of $N = 10$ sensors, versus 56 — quite a savings!

2.7.2 Filter Spawning. Filter spawning is a type of moving-bank filter structure which focuses on improved parameter identification with computational savings. This architecture was implemented to help determine the amount of degradation suffered by a failing or failed actuator/sensor. This work originated with Fisher [54, 55]; it features a permanent collection of filters that are always on-line and an additional set of filters that may be called upon to augment the standard set as necessary. These augmenting sets of filters are “spawned” only after a specific actuator surface is declared to have failed to some degree, i.e., a partial failure is detected. These spawned filters assist in determining the level of failure.

2.7.3 Optimal Parameter Discretization. Several *ad hoc* techniques designed to discretize the parameter set were discussed in Section 2.3.3.2. Sheldon [175, 176, 177] cast the previously *ad hoc* process of continuous variable parameter discretization as an optimization problem designed to improve either the state or parameter estimation or state control/regulation. His research delivered a *static* optimal parameter discretization since the parameter set was discretized off-line prior to running the algorithm; hence the bank of (steady-state) filters is static ⁵⁰

⁵⁰Miller [149] and Vasquez [198] jointly extended the static optimal parameter discretization to enable a bank of (time-varying) filters to be redeclared while on-line, i.e., the parameter set may be rediscritized on-line; this is a *dynamic* discretization method.

The basic question being addressed by this algorithm is: “If we are allowed to choose the K points in the parameter set, where should they be placed in order to yield ‘optimal’ MMAE performance?” The optimal MMAE performance was defined as that which minimized the average mean-squared estimation error of states or parameters⁵¹.

The key ingredient of this approach is the creation of the cost functional used to optimize the parameter discretization for state estimation:

$$C_{\mathbf{x}} \triangleq \frac{\int_{\mathbb{A}} E\{[\mathbf{x} - \hat{\mathbf{x}}]^T \mathbf{W}_{\mathbf{x}} [\mathbf{x} - \hat{\mathbf{x}}]\} d\mathbf{a}}{\int_{\mathbb{A}} d\mathbf{a}} \quad (2.88)$$

where \mathbb{A} denotes the admissible parameter set discussed in Equations (2.36) and (2.37), $\mathbf{W}_{\mathbf{x}}$ is the user-specified weighting matrix used to emphasize or deemphasize various states, and the denominator, $\int_{\mathbb{A}} d\mathbf{a} \triangleq \int_{\mathbb{A}_J} \cdots \int_{\mathbb{A}_2} \int_{\mathbb{A}_1} da_1 da_2 \cdots da_J$, normalizes the contribution of the parameter set. Similarly, the cost functional minimized to optimize the parameter estimate is:

$$C_{\mathbf{a}} \triangleq \frac{\int_{\mathbb{A}} E\{[\mathbf{a} - \hat{\mathbf{a}}]^T \mathbf{W}_{\mathbf{a}} [\mathbf{a} - \hat{\mathbf{a}}]\} d\mathbf{a}}{\int_{\mathbb{A}} d\mathbf{a}} \quad (2.89)$$

where the matrix $\mathbf{W}_{\mathbf{a}}$ is a user-specified weighting scheme.

Sheldon made several assumptions in order to keep the mathematics tractable: the structure of the system model is known except for a J -vector of parameters, \mathbf{a} , from the infinite set \mathbb{A} , the bank is composed of K constant-gain, steady-state filters, and the MMAE converges to the “best” model in the “Baram” sense [16, 15] with probability one. Given that there are sufficient conditions for the MMAE to converge in the Baram sense, Sheldon found that the MMAE will converge to the

⁵¹Sheldon [175, 177] also applied his methodology to optimize the average mean-squared regulation error performance of the controller as criterion for the optimization of the MMAC.

j th filter governed by [175, 177]

$$l_j(t_i) = \min_{k=1,\dots,K} l_k(t_i) \quad (2.90)$$

where $l_k(t_i)$ is defined as the proximity of the k th filter generated as

$$l_k(t_i) \equiv \ln|\mathbf{A}_k(t_i)| + \text{tr} \{ \mathbf{A}_k^{-1}(t_i) [\mathbf{\Omega}_k(t_i) + \mathbf{R}_t(t_i)] \} \quad (2.91)$$

where $\mathbf{A}_k(t_i)$ is the filter-computed residual covariance and $\mathbf{\Omega}_k(t_i)$ is a quadratically-scaled steady-state state-estimation error autocorrelation, and $\mathbf{R}_t(t_i)$ is the truth measurement covariance — see [175] for the details.

Finally, Sheldon devised a five-step (off-line) algorithm to approximate numerically and minimize the cost functional. This algorithm required (among other things) a truth model of the system, the number of filters, K , to be implemented in the estimator, and an initial sample set, $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_K\}$ to begin the minimization [175]. Additionally, Sheldon extended his work to account for the design practice of placing lower bounds on the filter probabilities, p_k , to avoid the estimator lock-up problem previously discussed in Section 2.4.6.

2.7.4 Inter-Residual Distance Feedback. The success of the MMAE depends on the “distinguishability” of the models used in the bank of Kalman filters. To determine which parameter value to use, there must be appreciable differences between the characteristics of the residuals for the “correct” model versus the other, mismatched, filters. In the limit as the residuals become indistinguishable, the adaptation process is totally incapacitated. For fast and reliable parameter identification, assuming one of the hypothesized filters models is based on the true parameter value, the residuals should be as distinct as possible [124, 130].

Inter-residual distance feedback (IRDF) [123, 124] provides for the on-line modification of the Kalman filters in the MMAE bank for the purpose of maintaining

the distinguishability of the elemental filters. The discrimination property of the MMAE is preserved by continually adjusting the Kalman filters to keep the predicted measurements from becoming too close with respect to some performance metric.

Modification of the elemental filters is achieved by de-tuning the filters through the modulation of either the dynamic driving noise covariance, \mathbf{Q}_{dk} , or the new information $\mathbf{s}_k = \mathbf{K}_k \mathbf{r}_k$ directly, where \mathbf{K}_k is the k th Kalman filter gain matrix and \mathbf{r}_k is the residual of the k th Kalman filter. The modulation process is governed by a scalar quantity computed from a distance measure between the residuals. Lund stressed the trade-off between discrimination and state tracking; tracking is the ability of the filter to predict the state $\mathbf{x}(t_i)$ and output $\mathbf{z}(t_i)$ given the measurement history \mathbf{Z}_{i-1} . The trade-off is between the desire for good tracking capabilities when one model matches the true system versus the desire for the residuals of the various filters to be distinct from each other, enabling fast and reliable model discrimination. If small Kalman filter gains are used to de-emphasize the impact of the measurement information, then the residuals of the various filters will tend to be more distant from each other. Thus, the elemental filters should *not* be tuned totally independently of the other filters [130] and the distance between the residuals should be large enough for inter-residual distinguishability [124].

With that in mind, Lund defined the *inter-residual difference measure*

$$J_{jk}(t_i) \triangleq \mathbf{r}_{jk}^T(t_i) \mathbf{\Gamma}_{jk} \mathbf{r}_{jk}(t_i), \quad \forall j \neq k \quad (2.92)$$

where the *inter-residual difference* is defined as

$$\mathbf{r}_{jk}(t_i) \triangleq \mathbf{r}_j(t_i) - \mathbf{r}_k(t_i) \quad (2.93)$$

with the j th Kalman filter residual $\mathbf{r}_j(t_i)$ and $\mathbf{\Gamma}_{jk}$ is a positive definite scaling matrix that is often diagonal. To maintain distinguishable residuals, we seek to keep $J_{jk}(t_i)$ above some threshold we'll call J_{jk}^0 by adjusting filter gains.

A general approach⁵² used to keep the inter-residual distance measure, $J_{jk}(t_i)$, above some specified limit J_{jk}^0 is to vary the dynamics noise covariance, \mathbf{Q}_{dk} , which by inspection of Equations (2.22) through (2.24) in Section 2.3.2 directly modifies the Kalman filter gains. We adjust \mathbf{Q}_{dk} with a modulating parameter η by

$$\mathbf{Q}'_{dk}(t_i) = \eta(t_i)\mathbf{Q}_{dk}(t_i), \forall k \in \{1, 2, \dots, K\} \quad (2.94)$$

where $\eta(t_i) \in [\eta_{\min}, 1]$. In order to maintain system stability, the lower bound must be nonnegative, i.e., $\eta_{\min} \geq 0$; Lund [124] and Miller [149] both discuss how one might smartly choose $\eta(t_i)$. Simulations are often used to help determine good values.

If the system models are linear, as they are in this research, then another way to impart change in the system is by modulating the new information⁵³ $\mathbf{s}_k(t_i) \triangleq \mathbf{K}_k(t_i)\mathbf{r}_k(t_i)$ as we did the dynamics noise

$$\mathbf{s}'_k(t_i) = \eta(t_i)\mathbf{s}_k(t_i), \forall k \in \{1, 2, \dots, K\} \quad (2.95)$$

The filter gains are now pre-computable and only the modulation parameter $\eta(t_i)$ is computed on-line. Furthermore, modulating $\mathbf{s}_k(t_i)$ versus $\mathbf{Q}_{dk}(t_i)$ leads to a faster adaptation because we don't have to wait for the filter error covariance transients to die before the corresponding filter gains are changed.

⁵²While Lund's [123, 124] work was with continuous-time systems with sampled measurements, we are assuming that we have a discrete-time system (or at least we are using an equivalent discrete-time model for a continuous system) and thus the rest of the development will follow Miller's work [149].

⁵³For a discussion on the statistics of the new information, see for example Section 5.4 of Maybeck [129].

2.7.5 Probability-Based Discretization Method. The IRDF, discussed in the previous section, worked to ensure stable probability flow in the filter bank by maintaining adequate distinguishability among the elemental filters in a relatively fixed bank. In this section, we introduce a technique aimed at enhanced “motion” and “sizing” of the filter bank for a moving-bank MMAE. Recall that the primary advantage of the moving bank is that potentially fewer filters are required to identify the parameter or system mode. The central concept employed by the probability-based discretization method (PBDM) [198, 200, 201] is to determine the parameter values for the elemental (Kalman) filters based on the calculation of the probability $\text{pr}(\chi_k^2 \leq T_{\chi^2})$, where the chi squared random variable, χ_k^2 , is defined as

$$\chi_k^2 = \mathbf{r}_k^T \mathbf{A}_k^{-1} \mathbf{r}_k \quad (2.96)$$

which is the same as the familiar likelihood quotient, L_k , previously defined in Equation (2.46), for threshold T_{χ^2} . In most cases, χ_k^2 , is a generalized chi squared random variable since the filter-computed residual covariance is only equal to the true residual covariance when the model matches the real world (truth) condition, i.e., only when $\mathbf{a}_k = \mathbf{a}_t$ will we have a chi squared random variable. When $\chi_k^2 \leq m$, where m is the length of the residual vector, then we see that filter model is a good match to the true condition, while $\chi_k^2 \gg m$ is an indication of a poor match with the real world condition. Thus, rather than using the ad hoc movement and resizing rules discussed earlier, this algorithm attempts to rediscretize the parameter space dynamically to achieve dynamic movement and sizing of the filter bank using the information already available to the algorithm.

2.7.6 Residual Correlation Kalman Filter Bank. Hanlon [71, 73] harnessed the power of the subliminal dither, previously discussed in Section 2.4.11, using a

new hypothesis conditional probability calculation⁵⁴ to maximize the observability of failed flight control actuators. Since the dither is a highly time-correlated (usually periodic) signal, it can be used to mark the filters that are based on a poor model with respect to the real world with a spike at the known dither frequency in the residual sequence PSD plot. While the standard MMAE does not exploit the whiteness of the residual sequence for a properly matched system model, Hanlon’s algorithm directly harnesses this knowledge. The purposeful dither appearing in the measurement is canceled by a dither in the predicted measurement which matches the real world condition, while an incorrectly predicted measurement will not cancel the dither present in the actual sensor output. Thus, the end result is that the elemental filter which matches the real world failure will have zero-mean white residuals, while the filters that don’t match the real world failure will pass the dither at the chosen frequency.

An algorithm which utilizes the purposeful dither is the residual correlation Kalman filter bank (RCKFB) technique; this method is based on a time-invariant system model for the elemental filters as shown in Equations (2.40) and (2.41). The residual sequence is assumed to be ergodic⁵⁵ so that the periodogram can be used to estimate the PSD [99]. The hypothesis conditional PDF employed by the RCKFB hypothesis testing algorithm is formed using different components for the residual vector (\mathbf{r}) and covariance matrix (\mathbf{A}). The measurement is replaced by the estimated PSD of the measurement residual ($\hat{\Psi}$), while the analog of the predicted measurement is the conditional mean of the PSD measurement residual ($\bar{\Psi}$). The residual is given by the difference of the estimated PSD and the conditional mean of the PSD evaluated at the known dither frequency: $\mathbf{r}_\psi = \hat{\Psi} - \bar{\Psi}$. The covariance

⁵⁴The standard hypothesis testing algorithm is described in Section 2.3.3.3 with Equations (2.31) through (2.46).

⁵⁵An ergodic sequence has the property that moments such as the mean and correlation can be computed based on the time average over a single sequence versus over a set of realizations — a statistical average. Ergodic sequences are a subset of stationary sequences.

matrix of the estimated PSD (\mathbf{A}_ψ) is used in place of the filter-computed residual covariance matrix.

The RCKFB employs the same basic testing algorithm but with a different hypothesis conditional PDF formed by exchanging a few terms, as indicated in the previous paragraph. With these modifications, it *is* an MMAE method to find the large PSD content at the *known* frequency of the dither. The drawback of the RCKFB is that it takes slightly longer to identify the failure, but the advantage is that the amplitude of the purposeful dither signal injected into the flight control actuators can be subliminal. Hence a combination of the standard MMAE and the MMAE with a RCKFB may prove useful.

2.7.7 Modified Multiple Model Adaptive Estimation. In traditional MMAE architectures, the designer has had to make a tradeoff decision between an optimal design for state estimation and a design concerned with accurate parameter estimation [130, 124]. Miller, however, developed the Modified MMAE (M³AE) architecture [149] that exploits the benefits of an MMAE designed for accurate parameter estimation and that performs at least as well in state estimation precision as an MMAE designed specifically for accurate state estimation. This architecture offers enhanced design flexibility in optimizing each estimator for its intended purpose. The MMAE portion of the M³AE is designed for parameter estimation. The elemental filters of the MMAE are designed and tuned such that the resulting hypotheses are as distinguishable as possible from each other. This increases the MMAE's ability to detect parameter changes in the system accurately. The estimated parameter from the the standard MMAE framework is then fed into a single Kalman filter designed to accept the parameter estimate. This filter is designed for accurate state estimation conditioned on the measurements and knowledge of the parameters provided by the parameter estimator. Miller [149, 148] found that the M³AE performed better than the standard MMAE when blending occurred between two or more elemental filters in the parameter estimator portion of the M³AE. Thus, the parameter estimate $\hat{\mathbf{a}}$

from the MMAE within the M³AE algorithm would be different from (and better than) any of the hypothesized \mathbf{a}_k values used as a basis for that MMAE’s elemental filters. To ensure that the parameter estimate is a good blend of two or more parameter, we need to have a fairly fine discretization of the parameter set so that a single elemental filter does not absorb all of the probability weight. The moving-bank MMAE design fits this scenario quite well and in fact, the M³AE doesn’t typically provide any substantial improvements in performance without a moving bank of filters.

2.8 *A Final Note*

The goal of this background chapter was to prepare the reader for the research in the subsequent chapters. We have satisfied that requirement, however, there are a plethora of advanced techniques and applications that we didn’t have space to discuss. The interested reader is encouraged to dive into the literature to search out those advanced techniques in order to extend the research discussed in this document.

III. Infinite-Dimensional Sampled-Data Kalman Filter

3.1 Introduction

The original intent of multiple model adaptive estimation (MMAE) [125] was to extend the applicability of the linear discrete-time Kalman filter [95]¹ to problems in which there was uncertainty in the strength of the dynamics noise model. Subsequent work extended the applicability of MMAE to the other parameters characterizing the structure of the model and the statistics of the noise models [108, 129, 196, 132] as reviewed in Chapter II. In this chapter we shall extend the linear sampled-data Kalman filter to allow an infinite-dimensional state space description [38, 39, 40] thereby creating the infinite-dimensional sampled-data Kalman filter (ISKF). The ISKF formulation will enable us to apply many of the tools of Kalman filtering previously applied to finite-dimensional, lumped parameter systems described by a vector stochastic differential equation (SDE) to systems best described with distributed parameters [163] or time-delayed measurements, using a stochastic *partial* differential equation (SPDE) [38, 87, 40] or a stochastic *retarded* differential equation (SRDE) [102, 38, 18, 40, 103].

It is well known that the Kalman filter optimally combines the state estimate from a static (sampled-data measurement update) minimum variance unbiased (MVU) estimator with state predictions based on a presumed dynamics model to estimate the state recursively [91, 129, 3, 29, 14]. The development of our ISKF extension of Kalman's filter [95] was primarily influenced by the presentations given by McGarty [141], Luenberger [122], Catlin [33], and Scharf [170], augmented by the infinite-dimensional linear systems theory reported by Curtain and Pritchard [38] and the more general framework of stochastic equations in infinite dimensions by

¹Kalman's seminal work [95] has been republished in collections edited by Sorenson [183] and Başar [12].

Da Prato and Zabczyk [40]. Additionally, most of the linear operator and transformation theory was gleaned from Naylor and Sell [154] as well as an introduction to topics in probability and measure theory which were more fully studied in Grigoriu [66] and Royden [166], respectively. A recent text by Bobrowski [24] combines the study of functional analysis, probability theory, and stochastic processes; his work aided this author in combining these concepts in a coherent fashion in the pages that follow.

The following section briefly discusses several topics in mathematical and functional analysis and probability theory. Most of the section is presented as definitions (with a few lemmas which we will prove since they are not common) and a few well-known theorems presented without proof. Next, we develop the new linear infinite-dimensional MVU estimator (LIMVUE) and then we generalize the LIMVUE to allow for recursive measurements. Then, we present an infinite-dimensional dynamics model which leads us to the new ISKF. We conclude this chapter with the generalized infinite-dimensional multiple model adaptive estimation (GIMMAE) framework.

3.2 Preliminary Concepts

We have assumed that the reader is familiar with the abstract mathematical concept of a vector space and in particular that of a linear vector space, hereafter referred to simply as a linear space. However, we will introduce some basic definitions and theorems to familiarize the reader with our notation and conventions. These results are well known and are thus given without proof. On the other hand, the proofs in the following sections flow more smoothly with the lemmas that we propose and prove in this section. The order in which we present the following concepts is simply as we need them, thus dependent concepts are presented after the independent concepts; e.g., an inner product and inner product space are defined before we talk about a Hilbert space.

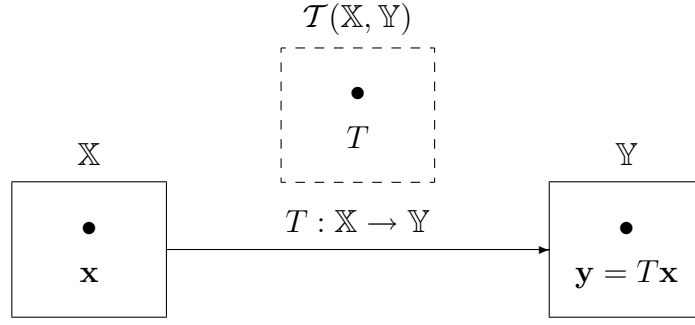


Figure 3.1 Boxology of a Transformation or Mapping.

In this dissertation, we often use a graphical depiction of the mappings employed in this research, referred to as the *boxology* of the mapping by Oxley [158]. These boxology diagrams help us to understand the relationships between the inputs, transformations, and outputs, and in particular, these diagrams identify the spaces in which the inputs, transformations, and outputs reside. This is in contrast to many control system and/or circuit diagrams that focus on the input/output or transfer function relationship itself. Figure 3.1 illustrates a simple example of the boxology employed in this dissertation and the following definition explains the notation.

Definition 1 (Transformations and Operators) *Let \mathbb{X} and \mathbb{Y} be vector spaces over the same field (\mathbb{R} or \mathbb{C}). Let $T : \mathbb{X} \rightarrow \mathbb{Y}$ represent a mapping from vector space \mathbb{X} to \mathbb{Y} , then T is a transformation and $\mathcal{T}(\mathbb{X}, \mathbb{Y})$ denotes the set of transformations from \mathbb{X} to \mathbb{Y} , and $T \in \mathcal{T}(\mathbb{X}, \mathbb{Y})$; this space of transformations is graphically depicted using the dashed box in Figure 3.1. If $T : \mathbb{Y} \rightarrow \mathbb{Y}$ is a mapping from \mathbb{Y} to itself, then we write that $T \in \mathcal{O}(\mathbb{Y})$, i.e., T is an operator.*

Now we shall add the structure of linearity to our mappings in the following definitions.

Definition 2 (Linear Transformations and Operators) *Let \mathbb{X} and \mathbb{Y} be vector spaces over the same field (\mathbb{R} or \mathbb{C}) and let α be a scalar. Let $L : \mathbb{X} \rightarrow \mathbb{Y}$ represent a mapping from vector space \mathbb{X} to \mathbb{Y} . L is a linear transformation if [154]:*

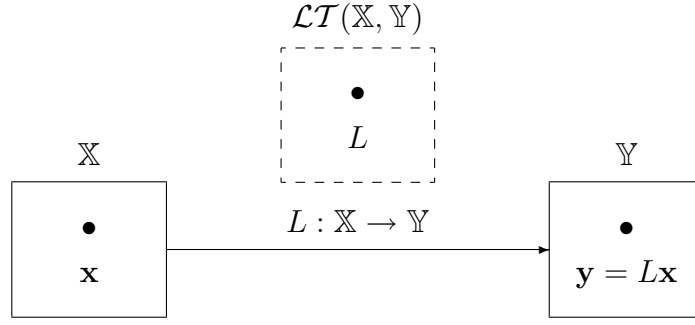


Figure 3.2 Boxology of a Linear Transformation.

1. *Scalar multiplication:* $L(\alpha\mathbf{x}) = \alpha L(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{X}$ and all scalars α , and
2. *Vector addition:* $L(\mathbf{x}_1 + \mathbf{x}_2) = L(\mathbf{x}_1) + L(\mathbf{x}_2)$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{X}$.

Otherwise, L is a nonlinear transformation. Notationally, $\mathcal{LT}(\mathbb{X}, \mathbb{Y})$ denotes the set of linear transformations from \mathbb{X} to \mathbb{Y} , and $L \in \mathcal{LT}(\mathbb{X}, \mathbb{Y})$; this relationship is illustrated in Figure 3.2. If $L : \mathbb{Y} \rightarrow \mathbb{Y}$ is a linear mapping from \mathbb{Y} to itself, then we write that $L \in \mathcal{LO}(\mathbb{Y})$, i.e., L is a linear operator.

A linear functional is a special type of linear transformation that maps the vector space over a scalar field to that scalar field, typically, the real numbers, \mathbb{R} . We use several linear functionals in this research; the inner product, defined later in this section, is one such example of a linear functional.

Definition 3 (Linear Functional) *Let \mathbb{Y} be a linear space over \mathbb{R} or \mathbb{C} . A linear functional over either field (\mathbb{R} or \mathbb{C}) simply maps (or transforms) a linear space to that field (\mathbb{R} or \mathbb{C}). [154]. We often denote the linear functional over \mathbb{R} using the familiar notation for a function: $\ell(\cdot)$, where the value $\ell(\mathbf{y}) \in \mathbb{R}$, for any $\mathbf{y} \in \mathbb{Y}$ [158].*

The most general type of topological space that we will employ in this research is a metric space; for our purposes, a metric space is defined as follows:

Definition 4 (Metric Space) *The pair (\mathbb{X}, d) is a metric space, when \mathbb{X} is a set and d is a real-valued function, a “distance” metric, defined for $\mathbf{x}, \mathbf{y} \in \mathbb{X}$ which adheres to the following axioms [154]:*

1. *Positive:* $d(\mathbf{x}, \mathbf{y}) \geq 0$ and $d(\mathbf{x}, \mathbf{x}) = 0$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{X}$.
2. *Strictly positive:* If $d(\mathbf{x}, \mathbf{y}) = 0$, then $\mathbf{x} = \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{X}$.
3. *Symmetry:* $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{X}$.
4. *Triangle inequality:* $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{X}$.

Example. Banach and Hilbert spaces are examples of metric spaces with the function d being the appropriate norm (used as the metric) of the difference of two vectors.

The norm of the difference of element pairs in a vector space is an example of a metric. When the norm is paired with a linear vector space, it forms a normed linear space.

Definition 5 (Normed Linear Space) *The linear space \mathbb{Y} is a normed linear space if there is a real-valued function, $\|\cdot\|$, which maps each $\mathbf{y} \in \mathbb{Y}$ to a real number $\|\mathbf{y}\| \in \mathbb{R}$. The norm is a “distance” metric that must obey the following axioms [154]:*

1. *Nonnegativity:* $\|\mathbf{y}\| \geq 0$ for all $\mathbf{y} \in \mathbb{Y}$.
2. *Positive definiteness:* $\|\mathbf{y}\| = 0$ if and only if $\mathbf{y} = \mathbf{0}$.
3. *Triangle inequality:* $\|\mathbf{y}_1 + \mathbf{y}_2\| \leq \|\mathbf{y}_1\| + \|\mathbf{y}_2\|$ for each $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{Y}$.
4. *Scalar multiplication:* $\|\alpha\mathbf{y}\| = |\alpha| \|\mathbf{y}\|$ for all scalars $\alpha \in \mathbb{R}$ and each $\mathbf{y} \in \mathbb{Y}$.

A normed linear space is denoted by $(\mathbb{Y}, \|\cdot\|)$ or more simply as \mathbb{Y} when the associated norm is understood.

To avoid confusion when dealing with multiple normed linear spaces and their associated norms, we often add a subscript to the norm notation, in this case, a normed linear space is denoted by $(\mathbb{Y}, \|\cdot\|_{\mathbb{Y}})$.

Definition 6 (Bounded Linear Transformation) *Let $B : \mathbb{X} \rightarrow \mathbb{Y}$ be a linear transformation. When \mathbb{X} and \mathbb{Y} are normed linear spaces, then the linear transfor-*

mation B is bounded, if there is a real number $M \geq 0$, such that

$$\|B\mathbf{x}\|_{\mathbb{Y}} \leq M\|\mathbf{x}\|_{\mathbb{X}} \quad \text{for all } \mathbf{x} \in \mathbb{X} \quad (3.1)$$

Let $\mathcal{BLT}(\mathbb{X}, \mathbb{Y})$ denote the set of bounded linear transformations (BLTs) from \mathbb{X} to \mathbb{Y} ; thus, $B \in \mathcal{BLT}(\mathbb{X}, \mathbb{Y})$, i.e., B is an element of the set of BLTs from \mathbb{X} to \mathbb{Y} [154]. When B is an operator, i.e., when $\mathbb{Y} = \mathbb{X}$, then we use the notation $B \in \mathcal{BLO}(\mathbb{X})$ to denote that B is a bounded linear operator (BLO).

To combine vectors from different spaces, we employ the direct sum of the spaces to *add* the vector components so that the new vector will be unique [122, 154].

Definition 7 (Direct Sum) Let \mathbb{X} and \mathbb{Y} be linear spaces over the same scalar field. The direct sum of \mathbb{X} and \mathbb{Y} , denoted by $\mathbb{X} \oplus \mathbb{Y}$, forms a new linear space. The underlying set of $\mathbb{X} \oplus \mathbb{Y}$ is formed by the Cartesian product, $\mathbb{X} \times \mathbb{Y}$, of sets \mathbb{X} and \mathbb{Y} . A point in $\mathbb{X} \times \mathbb{Y}$, is an ordered pair (\mathbf{x}, \mathbf{y}) , where $\mathbf{x} \in \mathbb{X}$ and $\mathbf{y} \in \mathbb{Y}$. Four notable properties of the direct sum are [154]:

1. Vector addition: $(\mathbf{x}_1, \mathbf{y}_1) \overset{\mathbb{X} \times \mathbb{Y}}{+} (\mathbf{x}_2, \mathbf{y}_2) = (\mathbf{x}_1 \overset{\mathbb{X}}{+} \mathbf{x}_2, \mathbf{y}_1 \overset{\mathbb{Y}}{+} \mathbf{y}_2)$ for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{X}$ and $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{Y}$.
2. Scalar multiplication: $\alpha(\mathbf{x}, \mathbf{y}) = (\alpha\mathbf{x}, \alpha\mathbf{y})$, for any scalar α , and every $\mathbf{x} \in \mathbb{X}$, and $\mathbf{y} \in \mathbb{Y}$.
3. Origin: $(0, 0) \in \mathbb{X} \times \mathbb{Y}$.
4. Negative: $-(\mathbf{x}, \mathbf{y}) = (-\mathbf{x}, -\mathbf{y})$ for every $\mathbf{x} \in \mathbb{X}$, and $\mathbf{y} \in \mathbb{Y}$.

Now that we have a method for uniquely combining the vectors from multiple normed linear spaces, we shall ascribe several useful properties to the normed linear spaces considered in this research, beginning with what it means for a sequence of vectors in a normed linear space to converge.

Definition 8 (Convergence) Let $(\mathbb{Y}, \|\cdot\|)$ be a normed linear space. An infinite sequence of vectors $\{\mathbf{y}_1, \mathbf{y}_2, \dots\} \subset \mathbb{Y}$ is said to converge to a vector \mathbf{y} if the sequence

$\{||\mathbf{y} - \mathbf{y}_1||, ||\mathbf{y} - \mathbf{y}_2||, \dots\}$ of real numbers converges to zero [154]. That is, given a real number $\epsilon > 0$ there exists an integer $N \in \mathbb{N}$ such that

$$||\mathbf{y}_n - \mathbf{y}|| < \epsilon \quad (3.2)$$

for all $n > N$.

Definition 9 (Cauchy Sequence) Let $(\mathbb{Y}, || \cdot ||)$ be a normed linear space. A sequence of vectors $\{\mathbf{y}_1, \mathbf{y}_2, \dots\} \subset \mathbb{Y}$ is called a Cauchy sequence if the sequence $\{||\mathbf{y}_m - \mathbf{y}_n|| : m, n \in \mathbb{N}\}$ of real numbers converges to zero, that is, given a real number $\epsilon > 0$, there exists an integer $N \in \mathbb{N}$, where $\mathbb{N} = \{1, 2, \dots\}$ is the set of natural numbers, such that

$$||\mathbf{y}_m - \mathbf{y}_n|| < \epsilon \quad (3.3)$$

for all $m, n > N$ [154].

Remark Note that every convergent sequence in a normed linear space is a Cauchy sequence and that, in general, the contrary is not true.

Definition 10 (Completeness) The normed linear space \mathbb{Y} is complete if every Cauchy sequence in \mathbb{Y} is a convergent sequence in \mathbb{Y} [154].

When a normed linear space is complete, it is often referred to by the special name:

Definition 11 (Banach Space) A complete normed linear space is also called a Banach space [154].

Example. An important Banach space for our research is the space $\mathbb{L}_{[a,b]}^p$ of Lebesgue measurable functions² for $1 \leq p < \infty$ with the finite norm [154]:

$$||\mathbf{x}|| = \left[\int_a^b |\mathbf{x}(\rho)|^p d\rho \right]^{1/p} \quad (3.4)$$

²To be precise, the equivalence class of Lebesgue measurable functions [154].

The Lebesgue functions are those functions, \mathbf{x} , defined on a closed real interval $[a, b]$ such that

$$\int_a^b |\mathbf{x}(\rho)|^p d\rho < \infty \quad (3.5)$$

i.e., functions that are absolutely integrable (for $p = 1$), square integrable (for $p = 2$), etc.

We note that Banach spaces are important in optimization problems (such as Kalman filtering) because the property of being a Cauchy sequence provides us a way of determining if a sequence of vectors in the Banach space does, indeed, converge in the space even when the limit of the sequence is unknown [122]. Another important property is the geometrical structure of the inner product. With the addition of this structure, we define another linear vector space, the inner product space. In general, an inner product space is defined over a scalar field, \mathbb{F} . Usually, the field in question is either the complex numbers, \mathbb{C} , or the real numbers, \mathbb{R} . In this research, we focus on an inner product space over the reals, \mathbb{R} .

Definition 12 (Inner Product Space) *Let \mathbb{Y} be a linear space over \mathbb{R} . An inner product on \mathbb{Y} over \mathbb{R} is a mapping that associates to each ordered pair of vectors \mathbf{y}_1 and \mathbf{y}_2 a real-valued scalar, denoted by $\langle \mathbf{y}_1, \mathbf{y}_2 \rangle$ that satisfies the following properties for $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3 \in \mathbb{Y}$ [154]:*

1. *Additivity:* $\langle \mathbf{y}_1 + \mathbf{y}_2, \mathbf{y}_3 \rangle = \langle \mathbf{y}_1, \mathbf{y}_3 \rangle + \langle \mathbf{y}_2, \mathbf{y}_3 \rangle$.
2. *Homogeneity:* $\langle \alpha \mathbf{y}_1, \mathbf{y}_2 \rangle = \alpha \langle \mathbf{y}_1, \mathbf{y}_2 \rangle$ for every $\alpha \in \mathbb{R}$.
3. *Symmetry:* $\langle \mathbf{y}_1, \mathbf{y}_2 \rangle = \langle \mathbf{y}_2, \mathbf{y}_1 \rangle$.
4. *Positive definiteness:* $\langle \mathbf{y}_1, \mathbf{y}_1 \rangle > 0$, when $\mathbf{y}_1 \neq \mathbf{0}$.

A linear space \mathbb{Y} with an inner product $\langle \cdot, \cdot \rangle$, written as $(\mathbb{Y}, \langle \cdot, \cdot \rangle)$ is known as an inner product space.

Note that the inner product generates a norm: $\|\mathbf{y}\| \triangleq (\langle \mathbf{y}, \mathbf{y} \rangle)^{1/2}$ — is often called the *induced norm* on \mathbb{Y} . If we have an inner product space $(\mathbb{Y}, \langle \cdot, \cdot \rangle)$, we often

abuse the notation and simply write \mathbb{Y} for convenience. When it is uncertain upon which space an inner product is defined, we will subscript the right angled bracket with the name of the space, as in: $\langle \cdot, \cdot \rangle_{\mathbb{Y}}$.

Example. The Euclidean space of real-valued N -vectors, \mathbb{R}^N , with inner product defined as: $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{n=1}^N x_n y_n$, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, where the superscript T represents the transpose operation, is an inner product space.

From the geometrical point of view, the inner product is a tool for comparing the relative “directions” of two vectors. When two vectors are completely aligned, their inner product is maximized, while vectors that are maximally skewed are said to be perpendicular or orthogonal when their inner product equals zero.

Definition 13 (Orthogonality) *Let $(\mathbb{Y}, \langle \cdot, \cdot \rangle)$ be an inner product space over \mathbb{R} . The vectors \mathbf{y}_1 and \mathbf{y}_2 are said to be orthogonal if their inner product is zero, i.e., if $\langle \mathbf{y}_1, \mathbf{y}_2 \rangle = 0$; thus we write $\mathbf{y}_1 \perp \mathbf{y}_2$, where the perpendicular symbol \perp reflects the geometrical interpretation [154]. Additionally, if \mathbb{A} and \mathbb{B} are two specific subsets of \mathbb{Y} , and $\langle \mathbf{a}, \mathbf{b} \rangle = 0$ for every $\mathbf{a} \in \mathbb{A}$ and $\mathbf{b} \in \mathbb{B}$, then these sets are orthogonal and we denote this by $\mathbb{A} \perp \mathbb{B}$ [154]. Furthermore, if $(\mathbb{A}, \langle \cdot, \cdot \rangle)$ and $(\mathbb{B}, \langle \cdot, \cdot \rangle)$ are subspaces of $(\mathbb{Y}, \langle \cdot, \cdot \rangle)$ and all of the subsets of \mathbb{A} and \mathbb{B} are orthogonal, then we say that the spaces are orthogonal and we denote this relation by $\mathbb{A} \perp \mathbb{B}$ [154].*

While the inner product of two vectors (on an inner product space) produces a scalar, the *outer* product of these two vectors defines a linear transformation. Some authors [37] refer to the operator created by the outer product as a rank-one operator; this is analogous to the creation of a rank-one matrix by the outer product of finite-dimensional vectors.

Definition 14 (Outer Product) *Let \mathbb{X} and \mathbb{Y} be inner product spaces. The outer product of two vectors $\mathbf{x} \in \mathbb{X}$ and $\mathbf{y} \in \mathbb{Y}$, denoted by $\mathbf{x} \diamond \mathbf{y}$, is defined in terms of the*

inner product on \mathbb{Y} as [38, 40]

$$(\mathbf{x} \diamond \mathbf{y})\mathbf{z} = \mathbf{x}\langle \mathbf{y}, \mathbf{z} \rangle, \text{ for every } \mathbf{z} \in \mathbb{Y} \quad (3.6)$$

The outer product forms a linear transformation, that is, $\mathbf{x} \diamond \mathbf{y} \in \mathcal{LT}(\mathbb{Y}, \mathbb{X})$ [40].

Example. Let x and y be continuous real-valued functions defined on the interval $[a, b]$ and denoted by $x, y \in C([a, b])$. The outer product is defined as: $(x \diamond y)z = x\langle y, z \rangle$ for all $z \in C([a, b])$. Thus, for all $z \in C([a, b])$,

$$[(x \diamond y)z](\rho) = x(\rho) \int_a^b y(\rho') z(\rho') d\rho' \quad (3.7)$$

The following technical lemma treats the case of an outer product when the second space is finite-dimensional:

Lemma 15 *Let \mathbb{X} and $\mathbb{Y} = \mathbb{R}^N$ be inner product spaces. For vectors $\mathbf{x} \in \mathbb{X}$ and $\mathbf{y} \in \mathbb{Y}$, the outer product can be represented as*

$$\mathbf{x} \diamond \mathbf{y} = \mathbf{x}\mathbf{y}^T \quad (3.8)$$

Proof of Lemma 15 We begin by using the definition of the outer product such that for every $\mathbf{z} \in \mathbb{Y}$ the following relation holds

$$(\mathbf{x} \diamond \mathbf{y})\mathbf{z} \triangleq \mathbf{x}\langle \mathbf{y}, \mathbf{z} \rangle \quad (3.9)$$

$$= \mathbf{x}(\mathbf{y}^T \mathbf{z}) \quad (3.10)$$

$$= (\mathbf{x}\mathbf{y}^T)\mathbf{z} \quad (3.11)$$

Therefore, Equation (3.8) follows since Equation (3.11) holds for all $\mathbf{z} \in \mathbb{Y}$. ■

The next lemma shows that the outer product has both the right and left distributive property.

Lemma 16 *Let \mathbb{X} and \mathbb{Y} be inner product spaces. (i). For vectors $\mathbf{x} \in \mathbb{X}$ and $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{Y}$, the right distributive property for the outer product is*

$$\mathbf{x} \diamond (\mathbf{y}_1 + \mathbf{y}_2) = (\mathbf{x} \diamond \mathbf{y}_1) + (\mathbf{x} \diamond \mathbf{y}_2) \quad (3.12)$$

(ii). *For vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{X}$ and $\mathbf{y} \in \mathbb{Y}$, the left distributive property for the outer product is*

$$(\mathbf{x}_1 + \mathbf{x}_2) \diamond \mathbf{y} = (\mathbf{x}_1 \diamond \mathbf{y}) + (\mathbf{x}_2 \diamond \mathbf{y}) \quad (3.13)$$

Remark Vector outer product takes precedence over vector addition; hence the parentheses on the right-hand sides of Equations (3.12) and (3.13), while unnecessary, are included for added clarity.

Proof of Lemma 16 (i). Begin by using the definition of the outer product for arbitrary $\mathbf{x} \in \mathbb{X}$ and $\mathbf{y}_1, \mathbf{y}_2, \mathbf{z} \in \mathbb{Y}$.

$$[\mathbf{x} \diamond (\mathbf{y}_1 + \mathbf{y}_2)]\mathbf{z} = \mathbf{x}\langle \mathbf{y}_1 + \mathbf{y}_2, \mathbf{z} \rangle \quad (3.14)$$

$$= \mathbf{x}(\langle \mathbf{y}_1, \mathbf{z} \rangle + \langle \mathbf{y}_2, \mathbf{z} \rangle) \quad (3.15)$$

$$= \mathbf{x}\langle \mathbf{y}_1, \mathbf{z} \rangle + \mathbf{x}\langle \mathbf{y}_2, \mathbf{z} \rangle \quad (3.16)$$

Using the definition of the outer product on the right-hand side of the equation again yields

$$[\mathbf{x} \diamond (\mathbf{y}_1 + \mathbf{y}_2)]\mathbf{z} = (\mathbf{x} \diamond \mathbf{y}_1)\mathbf{z} + (\mathbf{x} \diamond \mathbf{y}_2)\mathbf{z} \quad (3.17)$$

$$= [(\mathbf{x} \diamond \mathbf{y}_1) + (\mathbf{x} \diamond \mathbf{y}_2)]\mathbf{z} \quad (3.18)$$

where the last equality follows because we have two linear transformations applied to the same vector. Since Equation (3.18) holds for all $\mathbf{z} \in \mathbb{Y}$, then we get

$$\mathbf{x} \diamond (\mathbf{y}_1 + \mathbf{y}_2) = (\mathbf{x} \diamond \mathbf{y}_1) + (\mathbf{x} \diamond \mathbf{y}_2) \quad (3.12)$$

Thus property (i) of Lemma 16 holds.

(ii). Begin by using the definition of the outer product for arbitrary $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{X}$ and $\mathbf{y}, \mathbf{z} \in \mathbb{Y}$.

$$[(\mathbf{x}_1 + \mathbf{x}_2) \diamond \mathbf{y}] \mathbf{z} = (\mathbf{x}_1 + \mathbf{x}_2) \langle \mathbf{y}, \mathbf{z} \rangle \quad (3.19)$$

$$= \mathbf{x}_1 \langle \mathbf{y}, \mathbf{z} \rangle + \mathbf{x}_2 \langle \mathbf{y}, \mathbf{z} \rangle \quad (3.20)$$

Then we use the definition of the outer product again to get

$$[(\mathbf{x}_1 + \mathbf{x}_2) \diamond \mathbf{y}] \mathbf{z} = (\mathbf{x}_1 \diamond \mathbf{y}) \mathbf{z} + (\mathbf{x}_2 \diamond \mathbf{y}) \mathbf{z} \quad (3.21)$$

$$= [(\mathbf{x}_1 \diamond \mathbf{y}) + (\mathbf{x}_2 \diamond \mathbf{y})] \mathbf{z} \quad (3.22)$$

where once again, the last equality follows because we have two linear transformations applied to the same vector. Since Equation (3.22) holds for all $\mathbf{z} \in \mathbb{Y}$, then

$$(\mathbf{x}_1 + \mathbf{x}_2) \diamond \mathbf{y} = (\mathbf{x}_1 \diamond \mathbf{y}) + (\mathbf{x}_2 \diamond \mathbf{y}) \quad (3.13)$$

Thus property (ii) of the Lemma 16 holds. ■

For the next lemma, we will first introduce the concept of an adjoint operator.

Definition 17 (Adjoint) *Let \mathbb{X} and \mathbb{Y} be inner product spaces over the same field. If $B \in \mathcal{BLT}(\mathbb{X}, \mathbb{Y})$, then the adjoint of B , denoted B^* , is also a BLT, $B^* : \mathbb{Y} \rightarrow \mathbb{X}$, defined on \mathbb{X} and \mathbb{Y} such that the following holds true for all $B^* \in \mathcal{BLT}(\mathbb{Y}, \mathbb{X})$, $\mathbf{x} \in \mathbb{X}$, and $\mathbf{y} \in \mathbb{Y}$*

$$\langle B\mathbf{x}, \mathbf{y} \rangle_{\mathbb{Y}} = \langle \mathbf{x}, B^*\mathbf{y} \rangle_{\mathbb{X}} \quad (3.23)$$

Remark The adjoint of B satisfies the property: $\|B^*\| = \|B\|$ [154].

Lemma 18 *Let \mathbb{U} , \mathbb{V} , \mathbb{X} , and \mathbb{Y} be inner product spaces over the same field. For vectors $\mathbf{x} \in \mathbb{X}$ and $\mathbf{y} \in \mathbb{Y}$, and linear transformations $A \in \mathcal{LT}(\mathbb{X}, \mathbb{U})$ and $B \in$*

$\mathcal{LT}(\mathbb{Y}, \mathbb{V}),$

$$(A\mathbf{x}) \diamond (B\mathbf{y}) = A(\mathbf{x} \diamond \mathbf{y})B^* \quad (3.24)$$

where B^* is the adjoint of B .

Proof of Lemma 18 The outer product $(A\mathbf{x}) \diamond (B\mathbf{y})$ is defined by

$$[(A\mathbf{x}) \diamond (B\mathbf{y})]\mathbf{z} = A\mathbf{x}\langle B\mathbf{y}, \mathbf{z} \rangle \text{ for all } \mathbf{z} \in \mathbb{V} \quad (3.25)$$

Note that $\langle B\mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{y}, B^*\mathbf{z} \rangle$. Thus, for all $\mathbf{z} \in \mathbb{V}$,

$$[(A\mathbf{x}) \diamond (B\mathbf{y})]\mathbf{z} = A\mathbf{x}\langle \mathbf{y}, B^*\mathbf{z} \rangle \quad (3.26)$$

$$= A(\mathbf{x} \diamond \mathbf{y})B^*\mathbf{z} \quad (3.27)$$

Since Equation (3.27) holds for every $\mathbf{z} \in \mathbb{V}$, we get

$$(A\mathbf{x}) \diamond (B\mathbf{y}) = A(\mathbf{x} \diamond \mathbf{y})B^* \quad (3.24)$$

Therefore, the lemma holds. ■

As with the normed linear space, when we add completeness to the topology of an inner product space, we give it a special name.

Definition 19 (Hilbert Space) A complete inner product space is also called a Hilbert space [154].

Example. An important Hilbert space that we employ in our research is the space of real-valued Lebesgue measurable and square integrable functions over the interval $[a, b]$, denoted by $\mathbb{L}_{[a,b]}^2$, as defined on page 3-7, with the inner product [122]

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{L}_{[a,b]}^2} = \int_a^b \mathbf{x}(\rho) \mathbf{y}(\rho) d\rho \quad (3.28)$$

Now that we have defined the Hilbert space, is there another topological property that we desire? Yes, we desire the topological property of separability. The separability of a metric space (which includes the Hilbert space) says that we can arbitrarily approximate any point in the space using a countable orthonormal basis [154]. Thus, we are interested in the separable Hilbert space so that we can arbitrarily approximate any point in the space using an orthonormal basis. But before we define the separable Hilbert space, we shall introduce a few other topics: the orthonormal set and the maximal orthonormal set.

Definition 20 (Orthonormal Set) *Let $(\mathbb{Y}, \langle \cdot, \cdot \rangle)$ be an inner product space. The set, $\{\mathbf{y}_i \in \mathbb{Y} : i \in \mathbb{N}\}$, is called an orthonormal set if every pair of elements satisfies $\langle \mathbf{y}_i, \mathbf{y}_j \rangle = \delta_{ij}$ for all $i, j \in \mathbb{N}$, where δ_{ij} is the Kronecker delta function [154].*

Definition 21 (Maximal Orthonormal Set) *Let $(\mathbb{Y}, \langle \cdot, \cdot \rangle)$ be an inner product space. An orthonormal set $\mathbb{B} = \{\beta_1, \beta_2, \dots\}$ is called a maximal orthonormal set if there is no unit vector $\mathbf{y} \in \mathbb{Y}$ such that $\mathbb{B} \cup \{\mathbf{y}\}$ is an orthonormal set [154].*

Definition 22 (Separable Hilbert Space) *Let $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ be a Hilbert space. If every orthonormal set (spanning set) is countable and there is a maximal orthonormal set, then $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ is called a separable Hilbert space.*

Using these definitions, we can now propose a useful tool for representing a vector as a series of vectors.

Theorem 23 (Fourier Series) *If the set $\{\beta_1, \beta_2, \dots\}$ is any maximal orthonormal set and $\mathbf{y} \in \mathbb{H}$, then \mathbf{y} can be expressed by the Fourier series as [154, 166]:*

$$\mathbf{y} = \sum_{n=1}^{\infty} \alpha_n \beta_n \quad (3.29)$$

where $\alpha_n = \langle \mathbf{y}, \beta_n \rangle$ is the n th Fourier series coefficient. Moreover, $\|\mathbf{y}\|^2 = \sum_{n=1}^{\infty} \alpha_n^2$.

Proof of Theorem 23 See Naylor and Sell [154].

The next five operator and/or transformation properties are defined to help us describe the types of operators that the infinite-dimensional trace operator may act upon to produce a useful scalar; these are the trace-class operators [36, 37, 165, 40]. The trace operator is of central importance in an identity used to relate the covariance or correlation (which is defined using an outer product) to a particular inner product. With that in mind, the first operator property that we discuss is a useful extension of a familiar symmetry property of complex square matrices. We say that \mathbf{M} is a Hermitian (symmetric) matrix if and only if $\mathbf{M}^* = \mathbf{M} \in \mathbb{C}^{N \times N}$, where $*$ is the conjugate transpose; thus, $m_{pq} = m_{qp}^*$ for every element m_{pq} of \mathbf{M} . A generalization to a Hilbert space follows.

Definition 24 (Self-Adjoint Operator) *Let B be a BLO on a Hilbert space. If $B^* = B$, then B is said to be a self-adjoint operator [154].*

A slightly weaker property for (possibly unbounded) operators is given by the following.

Definition 25 (Symmetric Operator) *Let L be a linear operator on a dense subspace in an inner product space $(\mathbb{Y}, \langle \cdot, \cdot \rangle)$, i.e., the domain, $\mathcal{D}(L)$, is a dense set in \mathbb{Y} . L is said to be a symmetric operator if for every $\mathbf{x}, \mathbf{y} \in \mathcal{D}(L)$, $\langle L\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, L\mathbf{y} \rangle$ [39].*

Remark If either the domain of the adjoint operator, $\mathcal{D}(L^*)$, satisfies $\mathcal{D}(L^*) = \mathcal{D}(L)$, or if L is bounded, then a symmetric operator is also self-adjoint [39].

The following lemma demonstrates that the outer product operator is symmetric under certain circumstances.

Lemma 26 *Let $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ be a real Hilbert space. For every vector $\mathbf{u} \in \mathbb{H}$, the outer product $\mathbf{u} \diamond \mathbf{u}$ is a symmetric linear operator.*

Proof of Lemma 26 We shall show that $L = \mathbf{u} \diamond \mathbf{u}$ obeys $\langle L\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, L\mathbf{y} \rangle$ for every $\mathbf{u} \in \mathbb{H}$. Let $\mathbf{x}, \mathbf{y} \in \mathbb{H}$

$$\langle (\mathbf{u} \diamond \mathbf{u})\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{u} \langle \mathbf{u}, \mathbf{x} \rangle, \mathbf{y} \rangle \quad (3.30)$$

$$= \langle \mathbf{u}, \mathbf{x} \rangle \langle \mathbf{u}, \mathbf{y} \rangle \quad (3.31)$$

$$= \langle \mathbf{u}, \mathbf{y} \rangle \langle \mathbf{u}, \mathbf{x} \rangle \quad (3.32)$$

where the first line employed the definition of the outer product. Now reinserting $\langle \mathbf{u}, \mathbf{y} \rangle$ back into $\langle \mathbf{u}, \mathbf{x} \rangle$ yields

$$\langle (\mathbf{u} \diamond \mathbf{u})\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{u} \langle \mathbf{u}, \mathbf{y} \rangle, \mathbf{x} \rangle \quad (3.33)$$

$$= \langle (\mathbf{u} \diamond \mathbf{u})\mathbf{y}, \mathbf{x} \rangle \quad (3.34)$$

$$= \langle \mathbf{x}, (\mathbf{u} \diamond \mathbf{u})\mathbf{y} \rangle \quad (3.35)$$

Thus, the outer product operator, $\mathbf{u} \diamond \mathbf{u}$ is a symmetric operator. ■

Just like symmetric square matrices (and the set of real numbers), operators may possess the property of positiveness.

Definition 27 (Positive Operator) *A self-adjoint BLO B on a Hilbert space \mathbb{H} is positive if*

$$\langle \mathbf{x}, B\mathbf{x} \rangle \geq 0 \quad (3.36)$$

for all $\mathbf{x} \in \mathbb{H}$; this is denoted by $B \geq 0$. Similarly, the operator is strictly positive if

$$\langle \mathbf{x}, B\mathbf{x} \rangle > 0 \quad (3.37)$$

for all $\mathbf{x} \in \mathbb{H}$; this is denoted by $B > 0$ [154].

Remark Note that the word “positive” is not used altogether consistently when applied to operators and matrices by many authors and in this research. For example, a real-valued matrix $\mathbf{D} \in \mathbb{R}^{M \times M}$ is termed *positive semi-definite* when the inner

product $\langle \mathbf{x}, \mathbf{D}\mathbf{x} \rangle$ is nonnegative for every vector $\mathbf{x} \in \mathbb{R}^M$ and *positive definite* when the inner product is also nonzero, for all nonzero vectors in \mathbb{R}^M [189], i.e., a strictly positive scalar.

The next two properties apply to transformations (as well as operators). In terms of its spectral properties, a compact transformation is almost as simple as a matrix [154]. Recall that the eigenvalues of a matrix describe its spectrum and that the trace of a matrix is the sum of those eigenvalues. Hence it seems reasonable that a transformation that is “like” a matrix would be “traceable.” Additionally, the set of nuclear transformations, which includes the covariance and correlation transformations, is a subset of compact transformations.

Definition 28 (Compact Transformation) *Let \mathbb{X} and \mathbb{Y} be two Banach spaces over the same field. Let $L : \mathbb{X} \rightarrow \mathbb{Y}$ be a linear transformation that maps \mathbb{X} to \mathbb{Y} . L is said to be a compact transformation if $L(\mathbb{D})$ lies in a compact (or closed) subset of \mathbb{Y} , where $\mathbb{D} = \{\mathbf{x} \in \mathbb{X} : \|\mathbf{x}\| \leq 1\}$ [154].*

Remark Compactness is also referred to as complete continuity of a linear transformation [43, 216].

Definition 29 (Nuclear Transformation) *Let \mathbb{X} and \mathbb{Y} be Banach spaces over the same field, \mathbb{X}^* be the dual space of \mathbb{X} , i.e., the space of continuous linear functionals defined on the space, and $\mathcal{BLT}(\mathbb{X}, \mathbb{Y})$ be a Banach space of BLTs from \mathbb{X} into \mathbb{Y} . A transformation $B \in \mathcal{BLT}(\mathbb{X}, \mathbb{Y})$ is said to be nuclear if there exist two sequences $\{\alpha_1, \alpha_2, \dots\} \subset \mathbb{X}^*$ and $\{\mathbf{y}_1, \mathbf{y}_2, \dots\} \subset \mathbb{Y}$ such that [165, 40]*

$$\sum_j \|\mathbf{y}_j\| \cdot \|\alpha_j\| < \infty \quad (3.38)$$

and B is defined by

$$B(\mathbf{x}) = \sum_j \mathbf{y}_j \alpha_j(\mathbf{x}), \quad \mathbf{x} \in \mathbb{X} \quad (3.39)$$

Remarks (1) The space of all nuclear transformations (NT), from \mathbb{X} to \mathbb{Y} , is a Banach space, denoted $\mathcal{NT}(\mathbb{X}, \mathbb{Y})$ with the following norm [40]

$$\|B\|_{\mathcal{NT}} = \inf \left\{ \sum_j \|\mathbf{y}_j\| \cdot \|\alpha_j\| : B(\mathbf{x}) = \sum_j \mathbf{y}_j \alpha_j(\mathbf{x}) \right\} \quad (3.40)$$

(2) Since all nuclear transformations are compact [216] and all compact transformations are bounded and linear [154], nuclear transformations are necessarily compact, bounded and linear.

Lemma 30 *Let $\mathbf{u} \in \mathbb{U}$ and $\mathbf{v}, \mathbf{w} \in \mathbb{V}$ be vectors in separable Hilbert spaces of Lebesgue \mathbb{L}^2 functions. The outer product $\mathbf{u} \diamond \mathbf{v}$ is a nuclear transformation from \mathbb{V} to \mathbb{U} .*

Proof of Lemma 30 From Definition 14, we know that the outer product $\mathbf{u} \diamond \mathbf{v}$ is defined by the relation $(\mathbf{u} \diamond \mathbf{v})\mathbf{w} = \mathbf{u}\langle \mathbf{v}, \mathbf{w} \rangle$ for every $\mathbf{w} \in \mathbb{V}$. This matches the form (rather trivially) for transformation B in Equation (3.39). Since we have chosen Hilbert spaces of Lebesgue \mathbb{L}^2 functions, $\|\mathbf{u}\| \cdot \langle \mathbf{v}, \mathbf{w} \rangle$ is finite since both terms are finite. Hence, the outer product from one separable Hilbert space of Lebesgue \mathbb{L}^2 functions to another creates a nuclear transformation. ■

The trace of a matrix is the sum of the diagonal elements [129]. A deeper look shows that the trace of the matrix is equal to the sum of the eigenvalues [189]. The only criterion placed on the matrix is that it be square. To extend this concept to a Hilbert space, we must add more constraints beyond the fact that a square matrix corresponds to an operator. The trace operator may only be applied to trace-class operators. Since nuclear operators are the primary trace-class operators of interest to us in this research, we will forego a more in-depth discussion of the trace-class operators, see for example, references [37, 165, 40], for an extensive development.

Definition 31 (Trace Operator) (i). *Let L be a self-adjoint operator [154], a compact positive operator [37], or a nuclear operator [40] on a Hilbert space \mathbb{H} ; then*

the trace of L is given by

$$\text{tr } L = \sum_n \langle L\beta_n, \beta_n \rangle \quad (3.41)$$

where $\{\beta_1, \beta_2, \dots\}$ is any orthonormal basis on \mathbb{H} .

(ii). If L is a compact self-adjoint positive operator [154] on \mathbb{H} , then the trace of L is given by

$$\text{tr } L = \sum_n \lambda_n \quad (3.42)$$

where $\{\lambda_1, \lambda_2, \dots\}$ is the set of eigenvalues of L .

Remark A positive BLO defined on a Banach space is nuclear if and only if the trace of the operator is finite [40].

The linear space of real N -vectors, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, with associated inner product defined as: $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ and outer product: $\mathbf{x} \diamond \mathbf{y} = \mathbf{x} \mathbf{y}^T$ is a separable Hilbert space. It is common knowledge that the trace of the outer product is equal to the inner product³: $\text{tr}(\mathbf{x} \mathbf{y}^T) = \mathbf{x}^T \mathbf{y}$. In the next lemma, we extend this trace operator property for the case of (possibly) infinite-dimensional vectors in a separable Hilbert space.

Lemma 32 Let $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space of Lebesgue \mathbb{L}^2 functions. The trace of the outer product, $\text{tr}(\mathbf{x} \diamond \mathbf{y})$, is equal to the inner product $\langle \mathbf{x}, \mathbf{y} \rangle$, for any $\mathbf{x}, \mathbf{y} \in \mathbb{H}$, i.e.

$$\text{tr}(\mathbf{x} \diamond \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle \quad (3.43)$$

Proof of Lemma 32 From Lemma 30, we may employ Definition 31, to obtain, for every $\mathbf{x}, \mathbf{y} \in \mathbb{H}$,

$$\text{tr}(\mathbf{x} \diamond \mathbf{y}) = \sum_n \langle (\mathbf{x} \diamond \mathbf{y})\beta_n, \beta_n \rangle \quad (3.44)$$

³This property of the trace is simply a specific case of a more general result [27]: $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$ for appropriately dimensioned matrices \mathbf{A} and \mathbf{B} .

where $\{\beta_1, \beta_2, \dots\}$ is any orthonormal basis. Using the definition of the outer product, we can rewrite Equation (3.44) as

$$\text{tr}(\mathbf{x} \diamond \mathbf{y}) = \sum_n \langle \mathbf{x} \langle \mathbf{y}, \beta_n \rangle, \beta_n \rangle \quad (3.45)$$

Factoring out the inner product yields

$$\text{tr}(\mathbf{x} \diamond \mathbf{y}) = \sum_n \langle \mathbf{x}, \beta_n \rangle \langle \mathbf{y}, \beta_n \rangle \quad (3.46)$$

$$= \sum_n \mathbf{x}_n \mathbf{y}_n \quad (3.47)$$

$$= \langle \mathbf{x}, \mathbf{y} \rangle \quad (3.48)$$

where we note that \mathbf{x} and \mathbf{y} can be decomposed using the same orthonormal basis $\{\beta_1, \beta_2, \dots\}$, so that $\mathbf{x}_n = \langle \mathbf{x}, \beta_n \rangle$ and $\mathbf{y}_n = \langle \mathbf{y}, \beta_n \rangle$, respectively. Finally, we recognize that the sum in the third line is just the inner product. ■

For finite-dimensional vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, we know that \mathbf{x} and \mathbf{y} are orthogonal, by definition, whenever their inner product is zero: $\mathbf{x}^T \mathbf{y} = 0$. Additionally, it is true that whenever their outer product is zero, i.e., $\mathbf{x} \mathbf{y}^T = \mathbf{0} \in \mathbb{R}^{N \times N}$, that their inner product is also zero ($\mathbf{x}^T \mathbf{y} = 0 \in \mathbb{R}$) and therefore \mathbf{x} and \mathbf{y} are orthogonal. Observe that, since the trace of the zero outer product is trivially zero, the inner product is necessarily zero as well. In the following theorem, we extend this notion of using an outer product of vectors to identify the (geometrical) orthogonality of vectors in a more general setting of a separable Hilbert space of Lebesgue \mathbb{L}^2 functions. Note that this theorem helps us to preposition — in a natural fashion — the indispensable concept of statistical orthogonality that will be discussed in Definition 49.

Theorem 33 (Orthogonal Vectors) *Let $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space of Lebesgue \mathbb{L}^2 functions. Any two vectors, $\mathbf{x}, \mathbf{y} \in \mathbb{H}$, are orthogonal, i.e., $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, whenever*

$$\mathbf{x} \diamond \mathbf{y} = \mathbf{0} \in \mathcal{LO}(\mathbb{H}) \quad (3.49)$$

Proof of Theorem 33 Applying the trace operator to both sides of Equation (3.49) yields

$$\text{tr}(\mathbf{x} \diamond \mathbf{y}) = \text{tr}(\mathbf{0}) \quad (3.50)$$

$$\Rightarrow \text{tr}(\langle \mathbf{x}, \mathbf{y} \rangle) = 0 \quad (3.51)$$

$$\Rightarrow \langle \mathbf{x}, \mathbf{y} \rangle = 0 \quad (3.52)$$

where the second line follows since the trace of the outer product (defined on this separable Hilbert space of Lebesgue \mathbb{L}^2 functions) is equal to the trace of the inner product per Lemma 32 and the third lines gives the obvious result that the trace of a scalar is that scalar. Thus $\mathbf{x} \diamond \mathbf{y} = \mathbf{0}$ implies that $\mathbf{x} \perp \mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{H}$. ■

Since the topological concept of a closed subspace is important to the Projection Theorem, we will define it before we proceed further.

Definition 34 (Closed Subspace) *If (\mathbb{X}, d) and (\mathbb{Y}, d) are metric spaces such that $\mathbb{X} \subset \mathbb{Y}$, then we call (\mathbb{X}, d) a subspace of (\mathbb{Y}, d) . The subspace (\mathbb{X}, d) is closed if and only if every convergent sequence of vectors $\{x_1, x_2, \dots\}$ in \mathbb{X} has its limit in \mathbb{X} .*

The following two theorems are well-known and are thus stated without proof. The first theorem is a prelude to the projection theorem and it establishes the uniqueness of a vector that produces an error vector that is orthogonal to the subspace of interest.

Theorem 35 *Let \mathbb{Y} be an inner product space (not necessarily complete), \mathbb{S} a subspace of \mathbb{Y} , and \mathbf{y} an arbitrary vector in \mathbb{Y} . If there is a vector $\mathbf{s}_0 \in \mathbb{S}$ such that $\|\mathbf{y} - \mathbf{s}_0\| \leq \|\mathbf{y} - \mathbf{s}\|$ for all $\mathbf{s} \in \mathbb{S}$, then \mathbf{s}_0 is unique. A necessary and sufficient condition that \mathbf{s}_0 be a unique minimizing vector in \mathbb{S} is that the error vector $\mathbf{y} - \mathbf{s}_0$ be orthogonal to \mathbb{S} [122].*

Theorem 36 (Classical Projection Theorem) *Let \mathbb{H} be a Hilbert space and \mathbb{S} a closed subspace of \mathbb{H} . Corresponding to any vector $\mathbf{y} \in \mathbb{H}$, there is a unique vector*

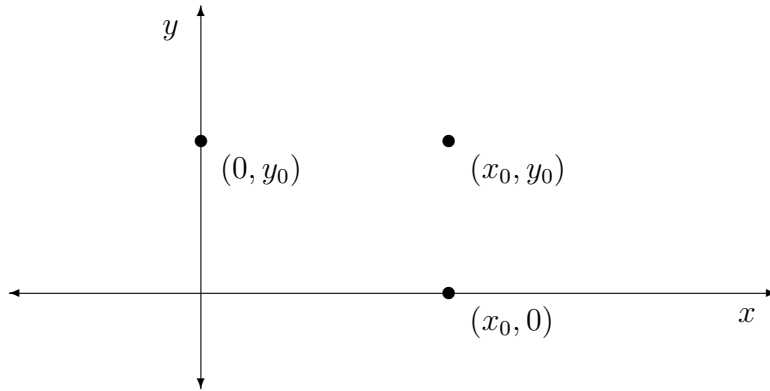


Figure 3.3 An Illustration of the Projection Theorem.

$\mathbf{s}_0 \in \mathbb{S}$ such that $\|\mathbf{y} - \mathbf{s}_0\| \leq \|\mathbf{y} - \mathbf{s}\|$ for all $\mathbf{s} \in \mathbb{S}$. Furthermore, a necessary and sufficient condition that $\mathbf{s}_0 \in \mathbb{S}$ be the unique minimizing vector is that the error vector $\mathbf{y} - \mathbf{s}_0$ be orthogonal to \mathbb{S} [122].

Example. Let $\mathbb{H} = \{(x, y) | x, y \in \mathbb{R}\}$ be the Euclidean 2-space (\mathbb{R}^2) with the usual inner product — the dot product. Let $\mathbb{S} = \{(x, y) | x, y \in \mathbb{R}, y = 0\}$ be the x -axis and $(x_s, 0)$ be an arbitrary vector (or a point in this case) in \mathbb{S} . We can easily visualize that any point in x - y plane, e.g. point (x_0, y_0) in Figure 3.3, corresponds to a unique point $(x_0, 0)$ on x -axis that is found by projecting onto the x -axis. The error vector between an arbitrary vector in \mathbb{H} and its corresponding vector in \mathbb{S} is given by: $(x_0, y_0) - (x_s, 0) = (x_0 - x_s, y_0 - 0) = (x_0 - x_s, y_0)$. We find x_s by minimizing the norm of the error; which in this example, is given by the square root of the sum of the squares: $[(x_0 - x_s)^2 + y_0^2]^{1/2}$. By inspection, we can see that the norm of the error is minimized when $x_s = x_0$. Thus, the optimal point on the x -axis, which minimizes the norm of the error is $(x_0, 0)$ as shown in Figure 3.3. Hence, the error vector is: $(x_0 - x_s, y_0)|_{x_s=x_0} = (0, y_0)$. Next, we use the dot product to show that the error vector, $(0, y_0)$, is orthogonal to the optimal projection of point (x_0, y_0) onto \mathbb{S} : $(x_0, 0)$. Therefore, $(0, y_0) \cdot (x_0, 0) = 0 \cdot x_0 + y_0 \cdot 0 = 0$. Q.E.D.

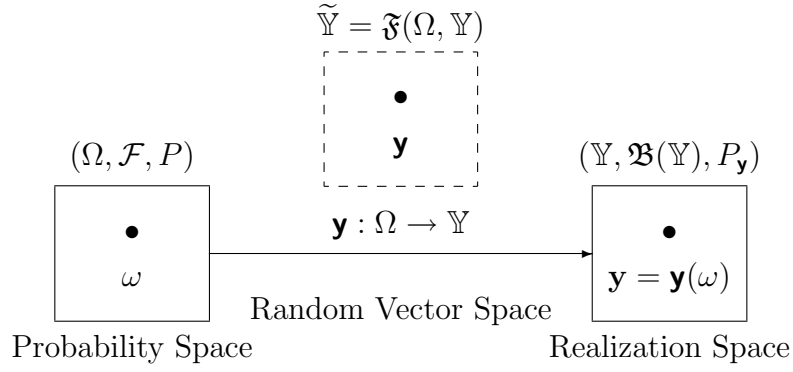


Figure 3.4 Boxology of a Random Vector.

For the static state estimation problem, we consider a Hilbert space of (possibly infinite-dimensional) random vectors⁴ $\mathbf{x} \in \tilde{\mathbb{X}} = \mathfrak{F}(\Omega, \mathbb{X})$ and a related linear space given by $\tilde{\mathbb{Z}} = \mathfrak{F}(\Omega, \mathbb{Z})$ for the single observation case. The notation $\mathfrak{F}(\cdot, \cdot)$ denotes a linear space of *functions*, Ω is a non-empty set called the sample space, and \mathbb{X} is the state space, i.e., the space of realizations of \mathbf{x} , and \mathbb{Z} is the observation space. Using the boxology previously employed, we can show graphically that for each experiment, an ω is “chosen” from the sample space Ω ; this choice dictates which \mathbb{Y} -valued random vector (or function) in $\tilde{\mathbb{Y}}$ represents the state as shown in Figure 3.4. Thus, Figure 3.4 illustrates how the probability space, random vector space, and realization space are interrelated. We use the tilde above \mathbb{Y} to help us associate the set of functions that map points (outcomes) in the sample space to vectors in the space of realizations. The remainder of the notation in the figure will be explained in the following definitions on the next few pages. Refer back to Figure 3.4 to see how the concepts are related.

Recall that our goal in this research is to solve an optimization problem in order to find the “best” state estimate, $\hat{\mathbf{x}}$, which minimizes the variance between the state estimate and the true state (i.e., the state estimation error) for a given measurement $\mathbf{z} \in \mathbb{Z}$. To accomplish this, we need to define the probability space, random

⁴In the language of vector spaces, every element in a vector space is called a vector. Thus, these random *vectors* may be random variables, random functions, or random matrices.

vector space, expected value, covariance, and conditional expectation, among other mathematical and probability theory topics. First, we will discuss a few technical topics from measure theory that lead up to a definition of the probability space.

Definition 37 (σ -Algebra and σ -Field) *Let \mathcal{S} be a nonempty set. A σ -algebra on \mathcal{S} is a collection of subsets of \mathcal{S} such that the necessary subsets in that collection are the set itself, the empty set, the complements of all the members in the collection and all countable unions of members. A σ -field is also called a σ -algebra [25].*

Definition 38 (Borel Sets) *For set \mathbb{A} , the collection $\mathfrak{B}(\mathbb{A})$ of Borel sets is the smallest σ -algebra which contains all of the open subsets of \mathbb{A} [166].*

Definition 39 (Measurable Space and Measurable Set) *A measurable space is a pair $(\mathbb{X}, \mathcal{G})$, consisting of nonempty set \mathbb{X} and σ -algebra \mathcal{G} of subsets of \mathbb{X} . A subset \mathbb{Y} of \mathbb{X} is called measurable (or measurable with respect to \mathcal{G}) if $\mathbb{Y} \in \mathcal{G}$. [166].*

Definition 40 (Measure and Measure Space) *Let \mathbb{X} be a nonempty set and \mathcal{G} be a σ -algebra defined on \mathbb{X} . A measure μ on the measurable space $(\mathbb{X}, \mathcal{G})$ is a nonnegative set function defined for all sets in \mathcal{G} and satisfying $\mu(\emptyset) = 0 \in \mathbb{R}$, where \emptyset is the empty or null set, and*

$$\mu \left(\bigcup_i \mathbb{E}_i \right) = \sum_i \mu(\mathbb{E}_i)$$

for any sequence $\mathbb{E}_1, \mathbb{E}_2, \dots$ of disjoint measurable sets. A measure space $(\mathbb{X}, \mathcal{G}, \mu)$ is a triplet formed by a measurable space $(\mathbb{X}, \mathcal{G})$ with a measure μ defined on \mathcal{G} . [166].

If the measure of \mathbb{X} is one, i.e., $\mu(\mathbb{X}) = 1$, and $0 \leq \mu(\mathbb{Y}) \leq 1$ for every \mathbb{Y} in \mathcal{G} , then μ may be called a probability measure. This leads us to define our probability space and other associated properties and terminology.

Definition 41 (Probability Space and Expectation) Suppose that the triplet (Ω, \mathcal{F}, P) is a complete probability space⁵, where Ω is a non-empty set called the sample space, \mathcal{F} is a σ -field which consists of a collection of subsets of Ω , called events, and P is a probability measure, a mapping that assigns a real number between zero and one (inclusive) to every event in \mathcal{F} , with the probability of the sure event $P(\Omega) = 1$. Let \mathbb{Y} be a Banach space. A \mathbb{Y} -valued random variable is a map $\mathbf{y} : \Omega \rightarrow \mathbb{Y}$ which is strongly measurable⁶ with respect to the probability measure P . If \mathbf{y} is integrable (in the sense of Bochner) on Ω , we define the expectation operator, E , with the integral expression [38]

$$E(\mathbf{y}) = \int_{\Omega} \mathbf{y} dP = \int_{\Omega} \mathbf{y}(\omega) P(d\omega) \quad (3.53)$$

(Note that \mathbf{y} is said to be Bochner integrable if $\int_{\Omega} \|\mathbf{y}\| dP < \infty$ [40].) The random vector \mathbf{y} induces a measure $P_{\mathbf{y}}$ on $\mathfrak{B}(\mathbb{Y})$, the Borel sets of \mathbb{Y} , which is defined as

$$P_{\mathbf{y}}(\mathbb{A}) = P\{\omega : \mathbf{y}(\omega) \in \mathbb{A}\} \quad (3.54)$$

for $\mathbb{A} \in \mathfrak{B}(\mathbb{Y})$, and thus $(\mathbb{Y}, \mathfrak{B}(\mathbb{Y}), P_{\mathbf{y}})$ is also a complete probability space. An equivalent way of expressing Equation (3.53) using the probability measure $P_{\mathbf{y}}$ is [33]

$$E(\mathbf{y}) = \int_{\mathbb{Y}} \mathbf{y} dP_{\mathbf{y}} \quad (3.55)$$

Remark The expectation operator is often subscripted with the pertinent random vector, as in $E_{\mathbf{y}}(\mathbf{y})$. Additionally, we often denote the mean of a random vector by $\mu_{\mathbf{y}} = E_{\mathbf{y}}(\mathbf{y})$.

⁵A probability space (Ω, \mathcal{F}, P) is complete if for every set $\mathbb{A} \subset \mathbb{B}$ such that $\mathbb{B} \in \mathcal{F}$ and $P(\mathbb{B}) = 0$ we have $\mathbb{A} \in \mathcal{F}$ so that $P(\mathbb{A}) = 0$ [66].

⁶Strong and weak measurability concepts are identical for separable Hilbert spaces [38].

Definition 42 (Joint Probability and Expectation) *Let the triplet (Ω, \mathcal{F}, P) be a complete probability space. Let \mathbb{X} and \mathbb{Y} be Banach spaces. The joint probability measure, $P_{\mathbf{x}, \mathbf{y}}$, induced by random vectors \mathbf{x} and \mathbf{y} on a collection of the events described by a relationship between the Borel sets $\mathfrak{B}(\mathbb{X})$ and $\mathfrak{B}(\mathbb{Y})$, is defined as*

$$P_{\mathbf{x}, \mathbf{y}}(\mathbb{A}, \mathbb{B}) = P\{\omega : \mathbf{x}(\omega) \in \mathbb{A} \text{ and } \mathbf{y}(\omega) \in \mathbb{B}\} \quad (3.56)$$

for $\mathbb{A} \in \mathfrak{B}(\mathbb{X})$ and $\mathbb{B} \in \mathfrak{B}(\mathbb{Y})$. The joint expectation of $g(\mathbf{x}, \mathbf{y})$ is then

$$E_{\mathbf{x}, \mathbf{y}}[g(\mathbf{x}, \mathbf{y})] = \int_{\Omega} g(\mathbf{x}, \mathbf{y}) dP = \int_{\mathbb{X}} \int_{\mathbb{Y}} g(\mathbf{x}, \mathbf{y}) dP_{\mathbf{x}, \mathbf{y}} \quad (3.57)$$

where g is a Baire function, i.e., a continuous function or the point-wise limit of a continuous function [129, 21].

Next, as a continuation of Lemma 18, given on page 3-13, we find the expected value of an outer product of random vectors.

Lemma 43 *Let $\tilde{\mathbb{U}}$, $\tilde{\mathbb{V}}$, $\tilde{\mathbb{X}}$, and $\tilde{\mathbb{Y}}$ be Hilbert spaces of random vectors. For random vectors $\mathbf{x} \in \tilde{\mathbb{X}}$ and $\mathbf{y} \in \tilde{\mathbb{Y}}$, and BLTs $A \in \mathcal{BLT}(\tilde{\mathbb{X}}, \tilde{\mathbb{U}})$ and $B \in \mathcal{BLT}(\tilde{\mathbb{Y}}, \tilde{\mathbb{V}})$,*

$$E_{\mathbf{x}, \mathbf{y}}[(A\mathbf{x}) \diamond (B\mathbf{y})] = A E_{\mathbf{x}, \mathbf{y}}(\mathbf{x} \diamond \mathbf{y}) B^* \quad (3.58)$$

where B^* is the adjoint of B and $E_{\mathbf{x}, \mathbf{y}}$ is the joint expectation operator.

Proof of Lemma 43 From Lemma 18, for random vectors we have

$$(A\mathbf{x}) \diamond (B\mathbf{y}) = A(\mathbf{x} \diamond \mathbf{y}) B^* \quad (3.59)$$

Taking the expectation of both sides yields

$$E_{\mathbf{x}, \mathbf{y}}[(A\mathbf{x}) \diamond (B\mathbf{y})] = E_{\mathbf{x}, \mathbf{y}}[A(\mathbf{x} \diamond \mathbf{y}) B^*] \quad (3.60)$$

Since A and B^* are nonrandom BLTs, the expectation commutes with A and B^* and we can pull them outside of the expectation, i.e., A and B^* are not a function of ω and thus they can be factored out of the integral defining the expectation in Equation (3.53) on page 3-25 and thus

$$E_{\mathbf{x}, \mathbf{y}}[(A\mathbf{x}) \diamond (B\mathbf{y})] = A E_{\mathbf{x}, \mathbf{y}}(\mathbf{x} \diamond \mathbf{y}) B^* \quad (3.58)$$

Therefore, the lemma holds. ■

The separable Hilbert spaces of Lebesgue \mathbb{L}^1 and \mathbb{L}^2 functions are the two most important spaces that we use to form our random vector spaces in this research. We give them here in the following example for $1 \leq p < \infty$.

Example. Let \mathbb{X} be a separable Hilbert space. The notation $\mathbb{L}^p(\Omega, P; \mathbb{X})$ denotes the separable Hilbert space of Lebesgue \mathbb{L}^p functions that are measurable with respect to P ; these functions map the sample space Ω to the realization space \mathbb{X} , hence \mathbf{x} is an \mathbb{X} -valued random vector. The \mathbb{L}^p norm is [154]

$$\|\mathbf{x}\|_{\mathbb{L}^p} = [E(\|\mathbf{x}\|_{\mathbb{X}}^p)]^{1/p} = \left[\int_{\Omega} \|\mathbf{x}(\omega)\|_{\mathbb{X}}^p dP(\omega) \right]^{1/p} \quad (3.61)$$

and $\|\mathbf{x}(\omega)\|_{\mathbb{X}}^p$ can be evaluated using Equation (3.4) when $\mathbb{X} = \mathbb{L}_{[a,b]}^p$. If \mathbb{X} is an N -dimensional Euclidian space \mathbb{R}^N and $p = 2$, then the *two norm* is written as: $\|\mathbf{x}(\omega)\|_{\mathbb{X}}^2 = \mathbf{x}^T(\omega) \mathbf{x}(\omega)$.

For finite-dimensional problems, we can refine our definition of the expectation operator in Equation (3.53) if we introduce the concept of a probability distribution function.

Definition 44 (Probability Distribution Function) *The probability distribution function for a $\mathbb{Y} = \mathbb{R}^N$ -valued random variable, $\mathbf{y} : \Omega \rightarrow \mathbb{Y}$, on the complete*

probability space (Ω, \mathcal{F}, P) , is given as [66]

$$F(\mathbf{y}) = P(\mathbf{y}^{-1}((-\infty, \mathbf{y}])) = P(\{\omega : \mathbf{y}(\omega) \leq \mathbf{y}\}) = P(\mathbf{y} \leq \mathbf{y}) \quad (3.62)$$

where the inequality is taken element by element for the case of a multi-dimensional vector. At the extremes, $F(-\infty) = 0$ and $F(\infty) = 1$.

Example. If we let $\mathbb{Y} = \mathbb{R}$, then Equation (3.62) is the accumulation of probability for $y(\omega)$ values less than or equal to y .

The following definitions lay out a series of properties pertaining to random vectors. For example, we are interested in the covariance operator — an extension of the familiar covariance matrix for infinite-dimensional systems. Since our work in this dissertation primarily uses separable Hilbert spaces (of random vectors), some of the following definitions and results apply only to separable Hilbert spaces.

Note that in the following pair of “second moment” definitions, we restrict our attention to a separable Hilbert space of Lebesgue \mathbb{L}^2 functions. Thus, the covariance and correlation operators (as well as the cross-covariance and cross-correlation transformations) will be bounded — and hence continuous — operators (transformations) since Lebesgue \mathbb{L}^2 functions are absolutely square integrable. Thus, while references [38, 40] stipulate that the covariance operator is symmetric, the covariance operator formed using Lebesgue \mathbb{L}^2 functions creates a bounded symmetric operator; hence the covariance operator is self-adjoint as noted in the remark following the definition of a symmetric operator.

Definition 45 (Covariance and Cross-Covariance) *Let \mathbb{X} be a separable Hilbert space. The covariance operator for an \mathbb{X} -valued random vector, $\mathbf{x} \in \tilde{\mathbb{X}} =$*

$\mathbb{L}^2(\Omega, P; \mathbb{X})$ is defined by [38, 40]

$$\Sigma(\mathbf{x}) \triangleq E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] \quad (3.63)$$

$$= \int_{\Omega} [(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] dP \quad (3.64)$$

$$= \int_{\mathbb{X}} [(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] dP_{\mathbf{x}} \quad (3.65)$$

where $\boldsymbol{\mu}_{\mathbf{x}} = E(\mathbf{x})$ and $P_{\mathbf{x}}$ is a probability measure induced by random vector \mathbf{x} . The covariance operator is self-adjoint, positive, and nuclear on \mathbb{X} .

Similarly, for random variables $\mathbf{x} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$ and $\mathbf{y} \in \tilde{\mathbb{Y}} = \mathbb{L}^2(\Omega, P; \mathbb{Y})$, the cross-covariance transformation is defined by

$$\Sigma(\mathbf{x}, \mathbf{y}) \triangleq E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})] \quad (3.66)$$

$$= \int_{\Omega} [(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})] dP \quad (3.67)$$

$$= \int_{\mathbb{X}} \int_{\mathbb{Y}} [(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})] dP_{\mathbf{x}, \mathbf{y}} \quad (3.68)$$

where $\boldsymbol{\mu}_{\mathbf{y}} = E(\mathbf{y})$ and $P_{\mathbf{x}, \mathbf{y}}$ is a probability measure induced jointly by random vectors \mathbf{x} and \mathbf{y} . Additionally, \mathbf{x} and \mathbf{y} are said to be uncorrelated whenever $\Sigma(\mathbf{x}, \mathbf{y}) = \mathbf{0}$.

Remarks (1) For the special case in which $\mathbf{x}, \mathbf{y} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$, the cross-covariance, while neither self-adjoint nor symmetric, is now a *nuclear operator* [40]. (2) Since the covariance operator is nuclear, it is also bounded and linear and is thus a BLO; hence, $\Sigma(\mathbf{x}) \in \mathcal{BLO}(\tilde{\mathbb{X}})$. Note that $(\mathcal{BLO}(\tilde{\mathbb{X}}), \|\cdot\|)$ is a Banach space, with the operator norm, provided that $\tilde{\mathbb{X}}$ is a complete normed linear space [154]. The cross-covariance transformation is also bounded and lives in a Banach space of BLTs, i.e., $\Sigma(\mathbf{x}, \mathbf{y}) \in \mathcal{BLT}(\tilde{\mathbb{Y}}, \tilde{\mathbb{X}})$.

Next, we will use the above definitions for the covariance operator and the cross-covariance transformation to define the non-central second moments: the correlation operator and the cross-correlation transformation.

Definition 46 (Correlation and Cross-Correlation) Let \mathbb{X} be a separable Hilbert space. The correlation operator for an \mathbb{X} -valued random vector, $\mathbf{x} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$, which is self-adjoint, positive, and nuclear, is defined by

$$\Xi(\mathbf{x}) \triangleq E(\mathbf{x} \diamond \mathbf{x}) \quad (3.69)$$

$$= \int_{\Omega} (\mathbf{x} \diamond \mathbf{x}) dP \quad (3.70)$$

$$= \int_{\mathbb{X}} (\mathbf{x} \diamond \mathbf{x}) dP_{\mathbf{x}} \quad (3.71)$$

where, as we saw before, $P_{\mathbf{x}}$ is a probability measure induced by random vector \mathbf{x} .

Similarly, for random variables $\mathbf{x} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$ and $\mathbf{y} \in \tilde{\mathbb{Y}} = \mathbb{L}^2(\Omega, P; \mathbb{Y})$, the cross-correlation transformation is defined by

$$\Xi(\mathbf{x}, \mathbf{y}) \triangleq E(\mathbf{x} \diamond \mathbf{y}) \quad (3.72)$$

$$= \int_{\Omega} (\mathbf{x} \diamond \mathbf{y}) dP \quad (3.73)$$

$$= \int_{\mathbb{X}} \int_{\mathbb{Y}} (\mathbf{x} \diamond \mathbf{y}) dP_{\mathbf{x}, \mathbf{y}} \quad (3.74)$$

where $P_{\mathbf{x}, \mathbf{y}}$ is a probability measure induced jointly by random vectors \mathbf{x} and \mathbf{y} .

Remarks (1) For the special case in which $\mathbf{x}, \mathbf{y} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$, the cross-correlation defined above, while neither self-adjoint nor symmetric, is now a *nuclear operator*. (2) Since the covariance operator is nuclear, it is also bounded and linear and is thus a BLO; thus it follows that the correlation operator is also a BLO, hence, $\Xi(\mathbf{x}) \in \mathcal{BLO}(\tilde{\mathbb{X}})$. The cross-correlation transformation is also bounded and lives in a Banach space of BLTs, i.e., $\Xi(\mathbf{x}, \mathbf{y}) \in \mathcal{BLT}(\tilde{\mathbb{Y}}, \tilde{\mathbb{X}})$.

An extremely useful identity for the covariance, $\Sigma(\mathbf{x})$, of random vector \mathbf{x} that we will use several times in this chapter is given next.

Lemma 47 *If $\mathbf{x} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$ is an \mathbb{X} -valued random vector for separable Hilbert space \mathbb{X} , then [40]*

$$\text{tr}[\Sigma(\mathbf{x})] = E\{\langle \mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}, \mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}} \rangle\} \quad (3.75)$$

where $\boldsymbol{\mu}_{\mathbf{x}}$ is the mean of random vector \mathbf{x} .

Proof of Lemma 47 Begin with the definition for the covariance and then by continuity interchange the linear operations for the trace [37] and expectation [154] to get

$$\text{tr}[\Sigma(\mathbf{x})] = \text{tr}\{E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})]\} \quad (3.76)$$

$$= E\{\text{tr}[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})]\} \quad (3.77)$$

$$= E\{\langle \mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}, \mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}} \rangle\} \quad (3.78)$$

where line three follows from Lemma 32. ■

As a corollary to Lemma 47, the identity can be extended for the cross-covariance operator $\Sigma(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}, \mathbf{y} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$.

Corollary 48 *If $\mathbf{x}, \mathbf{y} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$ is an \mathbb{X} -valued random vector on a separable Hilbert space, then*

$$\text{tr}[\Sigma(\mathbf{x}, \mathbf{y})] = E\{\langle \mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}, \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}} \rangle\} \quad (3.79)$$

where $\boldsymbol{\mu}_{\mathbf{x}}$ is the mean of random vector \mathbf{x} and $\boldsymbol{\mu}_{\mathbf{y}}$ is the mean of random vector \mathbf{y} .

For the special case of finite-dimensional random vectors, two vectors are termed statistically orthogonal whenever their cross-correlation matrix is the zero matrix [129], i.e., whenever $\Xi(\mathbf{x}, \mathbf{y}) \triangleq E(\mathbf{x}\mathbf{y}^T) = \mathbf{0}$. To extend this concept for infinite-dimensional systems, we propose the following definition:

Definition 49 (Statistically Orthogonal Vectors) *Let $\mathbf{x} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$ and $\mathbf{y} \in \tilde{\mathbb{Y}} = \mathbb{L}^2(\Omega, P; \mathbb{Y})$ be random vectors with separable Hilbert spaces \mathbb{X} and*

\mathbb{Y} , respectively. Any two vectors, $\mathbf{x} \in \tilde{\mathbb{X}}$ and $\mathbf{y} \in \tilde{\mathbb{Y}}$, are statistically orthogonal whenever their cross-correlation is the zero transformation, i.e.,

$$\Xi(\mathbf{x}, \mathbf{y}) \triangleq E(\mathbf{x} \diamond \mathbf{y}) = \mathbf{0} \in \mathcal{LT}(\mathbb{Y}, \mathbb{X}) \quad (3.80)$$

Remark If either \mathbf{x} or \mathbf{y} is a zero-mean random vector, then \mathbf{x} and \mathbf{y} are statistically orthogonal whenever \mathbf{x} and \mathbf{y} are either independent or uncorrelated.

A Hilbert space-valued random vector can be uniquely specified by its characteristic functional [38].

Definition 50 (Characteristic Functional) Consider a Hilbert space-valued random vector $\mathbf{x} \in \tilde{\mathbb{X}} = \mathbb{L}^1(\Omega, P; \mathbb{X})$, with the induced probability measure, $P_{\mathbf{x}}$. Define the mapping $\chi_{\mathbf{x}} : \mathbb{X} \rightarrow \mathbb{R}$ to be the characteristic functional for random vector \mathbf{x} and $j \triangleq \sqrt{-1}$ by [38, 40], where

$$\chi_{\mathbf{x}}(\boldsymbol{\xi}) \triangleq E_{\mathbf{x}}[\exp(j\langle \mathbf{x}, \boldsymbol{\xi} \rangle)] = \int_{\mathbb{X}} \exp(j\langle \mathbf{x}, \boldsymbol{\xi} \rangle) dP_{\mathbf{x}} \quad (3.81)$$

for all $\boldsymbol{\xi} \in \mathbb{X}$.

Later in our development of the ISKF, we will need to ascribe the Gaussian property to a random vector of interest, and subsequently to a stochastic process.

Definition 51 (Gaussian Random Vector, Gaussian Measure) The characteristic functional for a Hilbert space-valued random vector $\mathbf{x} \in \tilde{\mathbb{X}} = \mathbb{L}^1(\Omega, P; \mathbb{X})$, with Gaussian probability measure $P_{\mathbf{x}}$, is given by [38]

$$\chi_{\mathbf{x}}(\boldsymbol{\xi}) = \exp \left[j\langle \boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\xi} \rangle - \frac{1}{2} \langle \boldsymbol{\Sigma}_{\mathbf{x}} \boldsymbol{\xi}, \boldsymbol{\xi} \rangle \right] \quad (3.82)$$

where $\boldsymbol{\mu}_{\mathbf{x}} \in \mathbb{X}$ is the mean and the covariance operator $\boldsymbol{\Sigma}_{\mathbf{x}}$, with the modified notation, which was previously defined in Definition 45, is positive, self-adjoint, and nuclear.

There are several different types of convergence for random vectors: a sequence of random vectors can converge in probability, in mean square, with probability one, or in distribution [38]. If (Ω, \mathcal{F}, P) is a complete probability space and \mathbb{Y} is a Banach space, then convergence of a sequence of \mathbb{Y} -valued random vectors in $\mathbb{L}^2(\Omega, P; \mathbb{Y})$ is called mean square convergence [66]; it is defined next. The remaining forms of convergence will not be addressed further in this research.

Definition 52 (Mean Square Convergence) *Let \mathbb{Y} be a Banach space. A sequence $\{\mathbf{y}_1, \mathbf{y}_2, \dots\}$ of \mathbb{Y} -valued random vectors converges to \mathbf{y} in mean square sense if [38, 66]*

$$E(\|\mathbf{y}_n - \mathbf{y}\|_{\mathbb{Y}}^2) \rightarrow 0 \text{ as } n \rightarrow \infty \quad (3.83)$$

Definition 53 (Independence) *Let the triplet (Ω, \mathcal{F}, P) be a complete probability space.*

(i). *Events $A_1, A_2, \dots, A_n \in \mathcal{F}$ are independent if [45, 24]*

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i) \quad (3.84)$$

(ii). *Let $\mathbf{y}_i : \Omega \rightarrow \mathbb{Y}_i$ be a random vector with Borel set $\mathfrak{B}(\mathbb{Y}_i)$ for Banach space \mathbb{Y}_i . Random vectors $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$, are independent if [45]*

$$P\left(\bigcap_{i=1}^n \{\mathbf{y}_i \in \mathbb{A}_i\}\right) = \prod_{i=1}^n P(\mathbf{y}_i \in \mathbb{A}_i) \quad (3.85)$$

for all $\mathbb{A}_i \in \mathfrak{B}(\mathbb{Y}_i)$ for each $i = 1, 2, \dots, n$. Furthermore, both of these definitions can be extended to countably infinite number of objects, whether events or random vectors. A countably infinite number of objects forms an independent set if every finite subcollection is an independent set [45].

(iii). Let \mathbb{Y} be a Banach space. \mathbb{Y} -valued random vectors \mathbf{y}_1 and \mathbf{y}_2 are independent if $\{\omega : \mathbf{y}_1(\omega) \in \mathbb{A}_1\}$ and $\{\omega : \mathbf{y}_2(\omega) \in \mathbb{A}_2\}$ are independent sets in the σ -field \mathcal{F} for any Borel sets $\mathbb{A}_1, \mathbb{A}_2 \in \mathfrak{B}(\mathbb{Y})$ [38, 45].

(iv). Now let \mathbb{Y} be a separable Hilbert space. If \mathbf{y}_1 and \mathbf{y}_2 are in $L^1(\Omega, P; \mathbb{Y})$ and are independent, then they are also uncorrelated [38]:

$$E_{\mathbf{y}_1, \mathbf{y}_2}(\langle \mathbf{y}_1, \mathbf{y}_2 \rangle_{\mathbb{Y}}) = \langle E_{\mathbf{y}_1}(\mathbf{y}_1), E_{\mathbf{y}_2}(\mathbf{y}_2) \rangle_{\mathbb{Y}} \quad (3.86)$$

The MVU estimator, that we will develop in this chapter, is designed to exploit the statistical relationship between the observations and the states using the conditional expectation operator. The conditioning is accomplished relative to a sub σ -field of the observations σ -field.

Definition 54 (Sub σ -Field) Let the triplet (Ω, \mathcal{F}, P) be a probability space. Let \mathcal{S} be a subcollection of the σ -field \mathcal{F} , i.e., $\mathcal{S} \subset \mathcal{F}$. We call \mathcal{S} a sub σ -field if \mathcal{S} is also a σ -field of Ω [154].

Thus, we see that a sub σ -field \mathcal{S} may only give us partial (or incomplete) knowledge of the σ -field \mathcal{F} . If the sub σ -field \mathcal{S} is “equivalent” to the σ -field \mathcal{F} , except for a finite collection of sets of measure zero, then we gain complete knowledge, and our estimate using \mathcal{S} becomes as good as using \mathcal{F} since we will be taking an expectation of the state given the sure event.

Definition 55 (Conditional Expectation) Let the triplet (Ω, \mathcal{F}, P) be a complete probability space. Let \mathbb{X} be a Banach space. The conditional expectation of an \mathbb{X} -valued random vector \mathbf{x} relative to sub σ -field \mathcal{S} , denoted by $\mathcal{E}(\mathbf{x}|\mathcal{S})$, is defined by the relation [38, 154, 33, 40, 45]

$$\int_{\mathbb{A}} \mathbf{x} dP = \int_{\mathbb{A}} \mathcal{E}(\mathbf{x}|\mathcal{S}) dP \quad \text{for all } \mathbb{A} \in \mathcal{S} \quad (3.87)$$

Since this conditional expectation creates a random vector, we use a different symbol (\mathcal{E}) for the expectation operator. Note that $\mathcal{E}(\mathbf{x}|\mathcal{S})$ is uniquely defined by this relationship and must be measurable relative to the sub σ -field \mathcal{S} .

Remark When Equation (3.87) is evaluated for a specific event, $\mathbb{A}_1 \in \mathcal{S}$, then we write

$$E(\mathbf{x}|\mathbb{A}_1) = \int_{\mathbb{A}_1} \mathcal{E}(\mathbf{x}|\mathcal{S}) dP \quad (3.88)$$

Now we shall adapt this definition for conditional expectation for our eventual estimation purposes in the following example.

Example. Let the triplet (Ω, \mathcal{F}, P) be a complete probability space and let $(\mathbb{X}, \mathfrak{B}(\mathbb{X}), P_{\mathbf{x}})$ and $(\mathbb{Z}, \mathfrak{B}(\mathbb{Z}), P_{\mathbf{z}})$ be separable Hilbert spaces of \mathbb{X} - and \mathbb{Z} -valued random vectors, respectively. We condition the expectation on the Borel sets $\mathfrak{B}(\mathbb{Z})$, a sub- σ -field of \mathcal{F} , i.e., $\mathfrak{B}(\mathbb{Z}) \subset \mathcal{F}$, and then write Equation (3.87) as

$$\int_{\mathbb{A}} \mathbf{x} dP = \int_{\mathbb{A}} \mathcal{E}[\mathbf{x}|\mathfrak{B}(\mathbb{Z})] dP \quad \text{for all } \mathbb{A} \in \mathfrak{B}(\mathbb{Z}) \quad (3.89)$$

The boxology for this conditional expectation is shown in Figure 3.5. Note that $\tilde{\mathbb{Y}}$ is a subspace of $\tilde{\mathbb{X}}$ and the range of \mathbf{y} is a subset of \mathbb{X} , i.e., $\mathcal{R}(\mathbf{y}) \subset \mathbb{X}$; thus we would not expect the conditional mean estimator \mathbf{y} to produce an estimate of \mathbf{x} that is equal to $\mathbf{x} = \mathbf{x}(\omega)$ since \mathbf{y} maps vectors in \mathbb{Z} to a subspace of \mathbb{X} .

Consider the special case where \mathbf{x} is measurable relative to the σ -field $\mathfrak{B}(\mathbb{Z})$, sometimes written as $\mathbf{x} \in \mathfrak{B}(\mathbb{Z})$ [45], then $\mathcal{E}[\mathbf{x}|\mathfrak{B}(\mathbb{Z})] = \mathbf{x}$, i.e., \mathbf{x} is already the best guess for \mathbf{x} [45, 38]. At the other extreme, consider when \mathbf{x} is independent of $\mathfrak{B}(\mathbb{Z})$, then knowing $\mathfrak{B}(\mathbb{Z})$ does not change the expectation, hence $\mathcal{E}[\mathbf{x}|\mathfrak{B}(\mathbb{Z})] = E(\mathbf{x})$ [45].

We often write $\mathcal{E}(\mathbf{x}|\mathbf{z})$ in place of the rigorous notation $\mathcal{E}[\mathbf{x}|\mathfrak{B}(\mathbb{Z})]$. Additionally, we may add a subscript to the conditional expectation operator \mathcal{E} when needed for clarity, e.g., the conditional expectation of the sum of random vectors $\mathbf{x} + \mathbf{w}$ with respect to random vector \mathbf{x} given \mathbf{z} is written as $\mathcal{E}_{\mathbf{x}}(\mathbf{x} + \mathbf{w}|\mathbf{z})$, where \mathbf{w} is some

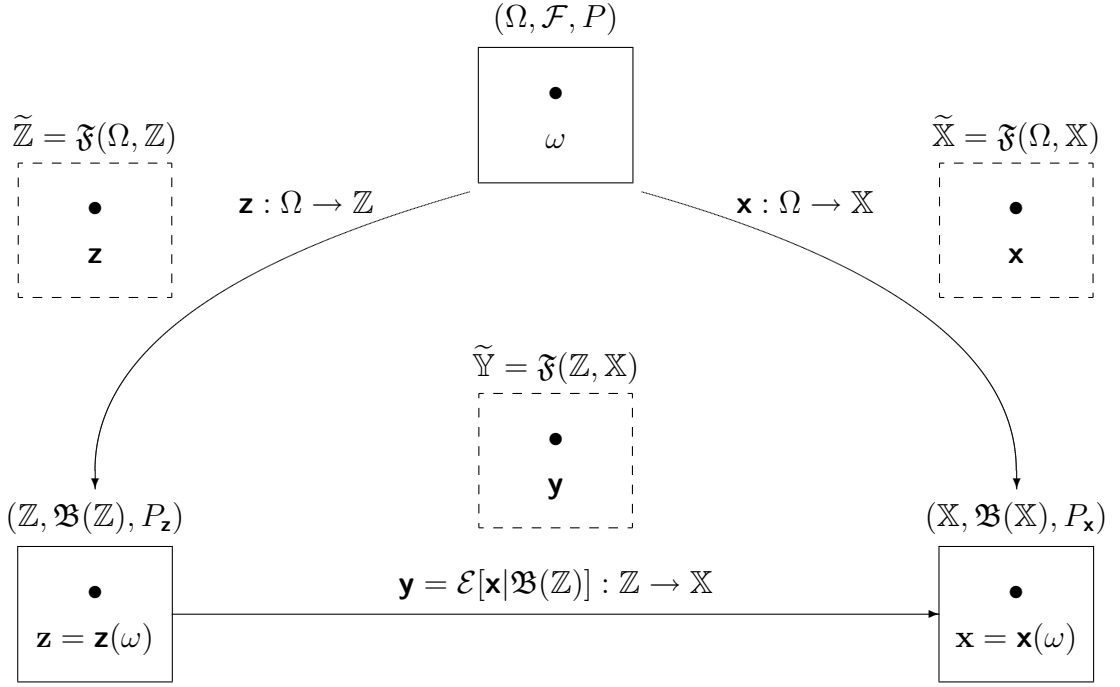


Figure 3.5 Boxology of a Conditional Expectation. Note that $\mathfrak{B}(\mathbb{Z}) \subset \mathcal{F}$ and $\tilde{\mathbb{Y}}$ is a subspace of $\tilde{\mathbb{X}}$.

\mathbb{W} -valued random vector. When the conditional expectation \mathcal{E} is evaluated for a particular event or realization of the sub σ -field or conditional random vector, e.g., $z \in \mathfrak{B}(\mathbb{Z})$ or $z(\omega) = z$, then we no longer create a random vector with the conditional expectation operation (as noted in the remark following the definition) and thus the notations $E\{x|z \in \mathfrak{B}(\mathbb{Z})\}$ and $E[x|z(\omega) = z]$ for the realizations of $\mathcal{E}\{x|\mathfrak{B}(\mathbb{Z})\}$ and $\mathcal{E}[x|z]$.

Definition 56 (Conditional Covariance) Let \mathbb{X} be a separable Hilbert space. The conditional covariance operator for an \mathbb{X} -valued random variable, $x \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$ relative to sub σ -field \mathcal{S} , denoted by $\tilde{\Sigma}(x|\mathcal{S})$, is given by

$$\tilde{\Sigma}(x|\mathcal{S}) \triangleq \mathcal{E}\{[x - \mathcal{E}(x|\mathcal{S})] \diamond [x - \mathcal{E}(x|\mathcal{S})]|\mathcal{S}\} \quad (3.90)$$

$$= \int_{\mathbb{A}} \{[x - \mathcal{E}(x|\mathcal{S})] \diamond [x - \mathcal{E}(x|\mathcal{S})]|\mathcal{S}\} dP \quad \text{for all } \mathbb{A} \in \mathcal{S} \quad (3.91)$$

Remark For a specific event, $\mathbb{A} \in \mathcal{S}$, the conditional covariance is no longer a random quantity; it is given by:

$$\Sigma(\mathbf{x}|\mathbb{A} \in \mathcal{S}) = \int_{\mathbb{A}} \{[\mathbf{x} - \mathcal{E}(\mathbf{x}|\mathcal{S})] \diamond [\mathbf{x} - \mathcal{E}(\mathbf{x}|\mathcal{S})]|\mathcal{S}\} dP \quad (3.92)$$

Since our random vectors are allowed to evolve or change over time, we must define what is meant by a stochastic process to account for the time-varying nature.

Definition 57 (Stochastic Process) *Let the triplet (Ω, \mathcal{F}, P) be a complete probability space. A stochastic process is a family of random vectors, given as $\{\mathbf{x}(t) : t \in \mathbb{T}\}$, that maps the product space $\mathbb{T} \times \Omega$ to the realization space \mathbb{X} for each fixed $t \in \mathbb{T}$, where \mathbb{T} is a set used to index the random vectors [141, 21]. The stochastic process is discrete if \mathbb{T} is countable and continuous if \mathbb{T} is homeomorphic to \mathbb{R} or some interval on \mathbb{R} . The stochastic process is denoted by either \mathbf{x} or $\mathbf{x}(\cdot, \cdot)$. Furthermore, both $\mathbf{x}(t)$ and $\mathbf{x}(t, \cdot)$ are used to denote a random vector, while $\mathbf{x}(t, \omega) = \mathbf{x}(t)$ is a realization of the random vector $\mathbf{x}(t, \cdot)$ and $\mathbf{x}(\cdot, \omega) = \mathbf{x}(\omega)$ is a sample of the stochastic process \mathbf{x} .*

Definition 58 (Covariance Kernel) *Let \mathbb{X} be a separable Hilbert space and $\mathbf{x}(t) \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$ be an \mathbb{X} -valued random vector for each $t \in \mathbb{T}$. The covariance kernel is an operator on a stochastic process $\{\mathbf{x}(t) : t \in \mathbb{T}\}$ defined by*

$$\Sigma[\mathbf{x}(t), \mathbf{x}(\tau)] \triangleq E\{[\mathbf{x}(t) - \boldsymbol{\mu}_{\mathbf{x}}(t)] \diamond [\mathbf{x}(\tau) - \boldsymbol{\mu}_{\mathbf{x}}(\tau)]\} \quad (3.93)$$

$$= \int_{\Omega} \{[\mathbf{x}(t) - \boldsymbol{\mu}_{\mathbf{x}}(t)] \diamond [\mathbf{x}(\tau) - \boldsymbol{\mu}_{\mathbf{x}}(\tau)]\} dP \quad (3.94)$$

$$= \int_{\mathbb{X}} \{[\mathbf{x}(t) - \boldsymbol{\mu}_{\mathbf{x}}(t)] \diamond [\mathbf{x}(\tau) - \boldsymbol{\mu}_{\mathbf{x}}(\tau)]\} dP_{\mathbf{x}(t)} \quad (3.95)$$

for $t, \tau \in \mathbb{T}$, where $\boldsymbol{\mu}_{\mathbf{x}}(t) = E[\mathbf{x}(t)]$, $\boldsymbol{\mu}_{\mathbf{x}}(\tau) = E[\mathbf{x}(\tau)]$, and $P_{\mathbf{x}(t)}$ is a probability measure induced by random vector $\mathbf{x}(t)$.

The models used by the estimators in this research can, in theory, be driven by most any noise process, provided that we know (at least) the first two moments. However, filter synthesis may prove exceedingly difficult (and the resulting filter suboptimal) unless the discrete-time model is driven by a Gaussian process. Note that the Wiener (or Brownian motion) process we use to drive the continuous-time model has independent increments that are Gaussian. The following definitions will describe Gaussian, independent increment, and finally, Wiener (or Brownian motion) processes.

Definition 59 (Gaussian Process) *Let the triplet (Ω, \mathcal{F}, P) be a complete probability space. An \mathbb{R} -valued stochastic process, $\{\mathbf{x}(t) : t \in \mathbb{T}\}$, is called a Gaussian process if every finite collection of random vectors is jointly Gaussian [42].*

Definition 60 (Independent Increment Process) *Let the triplet (Ω, \mathcal{F}, P) be a complete probability space and $\{\mathbf{x}(t) : t \in \mathbb{T}\}$ be a stochastic process, where \mathbb{T} is a discrete set such that $0 = t_0 < t_1 < \dots < t_f$. An increment is defined by the difference of two random vectors as $[\mathbf{x}(t_j) - \mathbf{x}(t_i)]$, where $t_i < t_j$ for $i < j$. If the disjoint increments $[\mathbf{x}(t_j) - \mathbf{x}(t_i)]$ and $[\mathbf{x}(t_l) - \mathbf{x}(t_k)]$ are independent, i.e., Definition 53 is satisfied for every disjoint pair of increments and without loss of generality, we have: $t_0 \leq t_i < t_j < t_k < t_l \leq t_f$, then the process is called an independent increment process.*

Many texts on probability theory, stochastic processes, and filtering theory devote an entire section (or chapter) to the Wiener (or Brownian motion⁷) process [42, 91, 141, 129, 21, 40, 66, 45, 24]. Falb defined a Wiener process in his development of the Kalman-Bucy filter on a Hilbert space [51]. We will follow Curtain and Pritchard's [38] abstract presentation in the following definition for a Wiener process or Brownian motion.

⁷The nineteenth century botanist Robert Brown studied the random thermal motion of grain particles suspended in a fluid and Norbert Wiener developed the mathematical foundation for this type of random motion [209, 21].

Definition 61 (Wiener Process or Brownian Motion) *Let the triplet (Ω, \mathcal{F}, P) be a complete probability space. The \mathbb{X} -valued stochastic process $\{\mathbf{b}(t) : t \in [0, t_f]\}$ is called a Wiener process (or Brownian motion)⁸ [38] on $[0, t_f]$ if it is an \mathbb{X} -valued process on $[0, t_f]$, such that $[\mathbf{b}(t) - \mathbf{b}(s)] \in \mathbb{L}^2(\Omega, P; \mathbb{X})$ for all $s, t \in [0, t_f]$ and*

1. $E[\mathbf{b}(t) - \mathbf{b}(s)] = \mathbf{0}$
2. $\Sigma[\mathbf{b}(t) - \mathbf{b}(s)] = (t - s)\mathbf{Q}$, where the constant $\mathbf{Q} \in \mathcal{BLO}(\widetilde{\mathbb{X}})$ and is symmetric, positive, and nuclear
3. $[\mathbf{b}(s_4) - \mathbf{b}(s_3)]$ and $[\mathbf{b}(s_2) - \mathbf{b}(s_1)]$ are independent whenever $0 \leq s_1 \leq s_2 \leq s_3 \leq s_4 \leq t_f$
4. $\mathbf{b}(t)$ has continuous sample paths on $[0, t_f]$

Additionally, the increment $[\mathbf{b}(t) - \mathbf{b}(s)]$ is Gaussian distributed with zero mean and covariance $\Sigma[\mathbf{b}(t) - \mathbf{b}(s)] = (t - s)\mathbf{Q}$, where the constant \mathbf{Q} is often called the diffusion of this constant-diffusion process featuring independent increments.

The Wiener process can be further generalized to include a time-varying diffusion [129]

$$\Sigma[\mathbf{b}(t) - \mathbf{b}(s)] = \int_s^t \mathbf{Q}(\tau) d\tau, \quad t > s \quad (3.96)$$

but we will not use this generality in the sequel since our problems of interest do not require this property.

Definition 62 (Discrete-Time White Noise Process) *Let the triplet (Ω, \mathcal{F}, P) be a complete probability space. A discrete-time white noise process, $\mathbf{n}(\cdot, \cdot)$, is defined as a collection of zero-mean independent random vectors $\{\mathbf{n}(t_i) : t_i \in \mathbb{T}\}$ with*

⁸We pay homage to Brown by denoting our Wiener process with the Arabic letter b, for Brownian motion.

covariance kernel operator

$$\Sigma[\mathbf{n}(t_i), \mathbf{n}(t_j)] = \begin{cases} \Sigma[\mathbf{n}(t_i)], & t_i = t_j \\ \mathbf{0}, & t_i \neq t_j \end{cases} \quad (3.97)$$

where $\Sigma[\mathbf{n}(t_i)]$ is positive (or positive semi-definite when the covariance takes the form of a matrix) and the cross-covariance between noise vectors at different times is zero since random vectors taken from a white noise process are independent in time and thus uncorrelated.

Remarks (1) Per Definition 45, the covariance operator is only guaranteed to be positive. (2) However, if we assume that there is “noise” in every dimension, i.e., that there are no zero eigenvalues, then the noise covariance is a strictly positive operator (or positive definite matrix). This “noise in every channel” assumption then gives rise to a definition that provides a sufficient condition for guaranteeing that the covariance operator (or matrix) for the measurements is invertible.

In the next section, we will begin to develop the tools needed to create the ISKF, but first, note that:

1. All of the random vectors in the remainder of the chapter are based on the fact that the triplet (Ω, \mathcal{F}, P) is a complete probability space as described in this section and specifically in Definition 41. For the sake of brevity, this statement will not be included in any of the definitions or theorems in the following sections, unless it is needed for clarity.
2. We have generally tried to avoid subscripting expectation operators to provide a cleaner look. The expectation is to be taken in a joint sense when there is more than one random vector involved and will be specifically noted when this is not the case. For example, we use $E(\mathbf{x})$ for $E_{\mathbf{x}}(\mathbf{x})$ and $E(\mathbf{x} + \mathbf{y})$ for $E_{\mathbf{x}, \mathbf{y}}(\mathbf{x} + \mathbf{y})$, while $E_{\mathbf{x}}(\mathbf{x} + \mathbf{y})$ will retain the subscripting.

3.3 Linear Infinite-Dimensional Minimum Variance Unbiased Estimator

There are two main classes of statistical signal processing techniques: the *Fisher* or *classical* approach, in which the signal (or parameter) of interest is assumed to be deterministic, yet unknown, and *Bayesian*, in which the signal (or parameter) of interest is random [100, 14]. Kalman applied the Bayesian approach to estimation theory [95]. Classical methods include techniques to produce least squares (LS) and maximum likelihood (ML) estimators. The Bayesian counterparts to these methods are the minimum mean-squared error (MMSE) and maximum a posteriori (MAP) estimators, respectively. An unbiased estimator that best minimizes the variance of the error is called the MVU estimator; this terminology is ordinarily associated with classical estimators [100]; however, some authors [33] use MVU as a synonym for MMSE estimators, and others [122] apply the terminology of MVU estimation to both classes of estimators. In this work, we will define the MVU estimator and employ it as a Bayesian estimation technique. A benefit of the Bayesian approach is that prior information is easily incorporated by the estimator, whereas classical techniques do not lend themselves as effortlessly to the admission of prior information. We can however, disguise the prior information in the form of a previous measurement output to achieve a similar effect. Luenberger [122] and Catlin [33] have derived MVU estimators for finite-dimensional models.

In this section, we give a series of definitions, lemmas, and theorems as we build up the machinery to derive several estimators. First, we create a linear infinite-dimensional MVU estimator (LIMVUE) for the case of correlated states and observations (CSO) and then follow with a more specific LIMVUE theorem to extend the applicability of finite-dimensional MVU estimators to infinite-dimensional problems using a technique often employed to find linear MMSE estimators. Before getting started with our LIMVUE theorems, we shall define several terms used in describing of Bayesian estimation. While there is more than one way to define an estimator, the following definition serves to support this research.

Definition 63 (Estimator, Estimate) Let $\mathbf{x} \in \tilde{\mathbb{X}}$ be the state and $\mathbf{z} \in \tilde{\mathbb{Z}}$ be a measurement. An estimator of random vector \mathbf{x} is defined as a random vector [33]

$$\hat{\mathbf{x}} = g \circ \mathbf{z} \quad (3.98)$$

where g is a Baire function and \circ is the composition operator. When we are given measurement $\mathbf{z}(\omega) = \mathbf{z}$, then an estimate of \mathbf{x} is

$$\hat{\mathbf{x}}(\omega) = g(\mathbf{z}(\omega)) = g(\mathbf{z}) \quad (3.99)$$

Thus, an estimate is a realization of the estimator.

Remark Scharf [170] and others call the state estimator $g \circ \mathbf{z}$ a *statistic*, i.e., a function of one or more random vectors that does not depend on any unknown parameters and a *sufficient statistic* if $g \circ \mathbf{z}$ carries all of the information about the data \mathbf{z} [86]. In our work, the statistic may well be a transformation applied to a function.

Definition 64 (Unbiased Estimate) Let $\hat{\mathbf{x}}$ be an estimator of random vector \mathbf{x} . Whenever $E_{\mathbf{x},\mathbf{z}}(\hat{\mathbf{x}}) = E_{\mathbf{x}}(\mathbf{x})$, we say that the estimator produces an unbiased estimate [14].

Theorem 65 (Conditional Mean Estimator) Let $(\tilde{\mathbb{X}}, \langle \cdot, \cdot \rangle_{\tilde{\mathbb{X}}})$ be a separable Hilbert space of \mathbb{X} -valued random vectors, and $\mathbf{x} \in \tilde{\mathbb{X}}$ be a random vector called the state. Let $(\tilde{\mathbb{Z}}, \langle \cdot, \cdot \rangle_{\tilde{\mathbb{Z}}})$ be a Hilbert space of \mathbb{Z} -valued random vectors, and $\mathbf{z} \in \tilde{\mathbb{Z}}$ be a random vector called the measurement. Then, the conditional (state) estimator is given by

$$\hat{\mathbf{x}} = \mathcal{E}(\mathbf{x}|\mathbf{z}) \quad (3.100)$$

and the error is then $[\mathbf{x} - \hat{\mathbf{x}}]$. The conditional estimator is endowed with the following properties:

1. The conditional state estimator $\hat{\mathbf{x}}$ is an unbiased estimator.
2. The error, $[\mathbf{x} - \hat{\mathbf{x}}]$, is orthogonal to the measurement data \mathbf{z} .
3. The error, $[\mathbf{x} - \hat{\mathbf{x}}]$, is orthogonal to the conditional estimator $\hat{\mathbf{x}}$.

Proof of Theorem 65 The conditional estimator can be found by correlating the state and the statistic [170]⁹:

$$E_{\mathbf{x},\mathbf{z}}\{\mathbf{x} \diamond (g \circ \mathbf{z})\} = E_{\mathbf{z}}\{\mathcal{E}_{\mathbf{x}}[\mathbf{x} \diamond (g \circ \mathbf{z})|\mathbf{z}]\} \quad (3.101)$$

$$= E_{\mathbf{z}}\{\mathcal{E}_{\mathbf{x}}(\mathbf{x}|\mathbf{z}) \diamond (g \circ \mathbf{z})\} \quad (3.102)$$

$$= E_{\mathbf{z}}\{\hat{\mathbf{x}} \diamond (g \circ \mathbf{z})\} \quad (3.103)$$

where $\hat{\mathbf{x}}$ is the conditional expectation

$$\hat{\mathbf{x}} = \mathcal{E}_{\mathbf{x}}(\mathbf{x}|\mathbf{z}) = \mathcal{E}(\mathbf{x}|\mathbf{z}) \quad (3.104)$$

which is a function of the data \mathbf{z} and it may or may not be a function of any unknown parameters. Equation (3.103) can be rewritten as

$$E_{\mathbf{x},\mathbf{z}}\{\mathbf{x} \diamond (g \circ \mathbf{z})\} - E_{\mathbf{z}}\{\hat{\mathbf{x}} \diamond (g \circ \mathbf{z})\} = \mathbf{0} \quad (3.105)$$

and since $E_{\mathbf{z}}\{\hat{\mathbf{x}} \diamond (g \circ \mathbf{z})\}$ is also $E_{\mathbf{x},\mathbf{z}}\{\hat{\mathbf{x}} \diamond (g \circ \mathbf{z})\}$ then we get

$$E\{[\mathbf{x} - \hat{\mathbf{x}}] \diamond (g \circ \mathbf{z})\} = \mathbf{0} \quad (3.106)$$

Thus, we see that the error, $\mathbf{x} - \hat{\mathbf{x}}$, is orthogonal to the statistic $g \circ \mathbf{z}$ per Definition 49 on page 3-31, since the cross-correlation is zero, i.e., $\Xi([\mathbf{x} - \hat{\mathbf{x}}], [g \circ \mathbf{z}]) = \mathbf{0}$.

⁹This proof is inspired by the finite-dimensional case in Scharf [170].

To prove property 1, we let the state estimator¹⁰ in Equation (3.106) be a constant *ones* vector, i.e., $g \circ \mathbf{z} = \mathbf{1}$, we obtain

$$E\{[\mathbf{x} - \hat{\mathbf{x}}] \diamond \mathbf{1}\} = E[\mathbf{x} - \hat{\mathbf{x}}] \diamond \mathbf{1} = \mathbf{0} \quad (3.107)$$

where $\mathbf{0}$ is the zero operator. This implies that $E[\mathbf{x} - \hat{\mathbf{x}}] = \mathbf{0}$, where $\mathbf{0}$ is now the zero vector. Thus $E(\mathbf{x}) = E(\hat{\mathbf{x}})$, hence the conditional estimator is unbiased.

Next, we show property 2. When the statistic is just the data, i.e., $g \circ \mathbf{z} = \mathbf{z}$, then we see that the error is orthogonal to the data

$$E\{[\mathbf{x} - \hat{\mathbf{x}}] \diamond \mathbf{z}\} = \mathbf{0} \quad (3.108)$$

Finally, for property 3, since $\hat{\mathbf{x}}$ is a function of only the data (a prerequisite for being a statistic), we let $g \circ \mathbf{z} = \hat{\mathbf{x}}$ to obtain

$$E\{[\mathbf{x} - \hat{\mathbf{x}}] \diamond \hat{\mathbf{x}}\} = \mathbf{0} \quad (3.109)$$

Thus the error is orthogonal to the conditional estimator. ■

Definition 66 (Minimum Variance Estimator) *For state $\mathbf{x} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$ and observation $\mathbf{z} \in \tilde{\mathbb{Z}} = \mathbb{L}^2(\Omega, P; \mathbb{Z})$, the estimator $\hat{\mathbf{x}} = g_o \circ \mathbf{z}$, given measurement \mathbf{z} , is called the minimum variance estimator if there exists an optimal Baire function g_o such that [33]*

$$\|\mathbf{x} - g_o \circ \mathbf{z}\| \leq \|\mathbf{x} - g \circ \mathbf{z}\| \quad (3.110)$$

¹⁰Note that $g \circ \mathbf{z}$ is not a statistic in this case since a ones vector is independent of the data.

holds for every Baire function g . Furthermore, the minimum variance estimator of \mathbf{x} based on measurement \mathbf{z} is given by¹¹

$$\hat{\mathbf{x}} = \mathcal{E}(\mathbf{x}|\mathbf{z}) \quad (3.111)$$

Remark Since \mathbb{L}^2 is a Hilbert space, the Projection Theorem (given in Theorems 35 and 36 on page 3-21) applies. Theorem 35 tells us that if g_o exists, then it is unique; hence the estimator $\hat{\mathbf{x}}$ is unique. Additionally, the error vector is orthogonal to every vector in the measurement space, i.e., $\langle \mathbf{x} - (g_o \circ \mathbf{z}), g \circ \mathbf{z} \rangle = 0$ for every Baire function g and every measurement $\mathbf{z} \in \tilde{\mathbb{Z}}$. The projection theorem in Theorem 36 says that a unique estimator exists for every measurement $\mathbf{z} \in \tilde{\mathbb{Z}}$.

Definition 67 (Linear Estimator) Let $\hat{\mathbf{x}}$ be an estimator of random vector \mathbf{x} and \mathbf{z} be a measurement used as an input to the estimator, $\hat{\mathbf{x}} = g \circ \mathbf{z}$ to produce an estimate, $\hat{\mathbf{x}}(\omega) = g(\mathbf{z}(\omega))$. The estimator, $\hat{\mathbf{x}}$ is said to be linear whenever $\mathbf{z} = \alpha_1 \mathbf{z}_1 + \alpha_2 \mathbf{z}_2$, for scalars α_1, α_2 , yields

$$g(\mathbf{z}(\omega)) = \alpha_1 g(\mathbf{z}_1(\omega)) + \alpha_2 g(\mathbf{z}_2(\omega)) \quad (3.112)$$

and nonlinear otherwise.

The Bayesian estimation technique rests on acquiring (either analytically or experimentally) the posteriori PDF, so that we can calculate the conditional moments of the state given the measurements. The next step involves picking the optimality criterion that will be used to produce the optimal estimator [31, 170, 100]. For our research, we chose to minimize the quadratic cost function, $C(\mathbf{e}) = \|\mathbf{e}\|^2$, where $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ is an estimation error for a cost function that places a high cost on large

¹¹Technically speaking, the conditional mean estimator proposed herein is actually the minimum variance estimator as the result of a theorem; see for example Catlin [33] for a proof that this is indeed the definition for an estimator that achieves the minimum variance.

errors and a small cost on small errors, to determine the optimal estimator. Thus, we shall find the estimator that minimizes the mean-squared error (MSE).

Definition 68 (Minimum mean-squared error estimator) *Let $(\mathbb{X}, \langle \cdot, \cdot \rangle_{\mathbb{X}})$ be a separable Hilbert space, $\mathbf{x} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$ be the random state vector, $(\mathbb{Z}, \langle \cdot, \cdot \rangle_{\mathbb{Z}})$ be a separable Hilbert space, and $\mathbf{z} \in \tilde{\mathbb{Z}} = \mathbb{L}^2(\Omega, P; \mathbb{Z})$ be the measurement vector. The MMSE estimator is the random vector which minimizes the Bayesian MSE between the estimator, $\hat{\mathbf{x}}_{\text{MMSE}}$, and the true state, \mathbf{x} , as [100, 170]*

$$\hat{\mathbf{x}}_{\text{MMSE}} = \arg \left\{ \min_{\hat{\mathbf{x}}} [E_{\mathbf{x}, \mathbf{z}} (||\mathbf{x} - \hat{\mathbf{x}}||_{\mathbb{X}}^2)] \right\} \quad (3.113)$$

Lemma 69 *The solution to Equation (3.113) is the conditional mean, $\hat{\mathbf{x}} = \mathcal{E}(\mathbf{x}|\mathbf{z})$.*

Remark Two important properties of this estimator are (1) it is unbiased, i.e., the estimator error is zero-mean, and (2) the estimator error and the measurements are uncorrelated, i.e., they are orthogonal [14].

Proof of Lemma 69 We begin with the definition of the MSE and then expand after adding a “smart” zero, $[\mathcal{E}(\mathbf{x}|\mathbf{z}) - \mathcal{E}(\mathbf{x}|\mathbf{z})]$ to obtain

$$\begin{aligned} E_{\mathbf{x}, \mathbf{z}} (||\mathbf{x} - \hat{\mathbf{x}}||_{\mathbb{X}}^2) \\ = E_{\mathbf{x}, \mathbf{z}} \{ ||\mathbf{x} + [\mathcal{E}(\mathbf{x}|\mathbf{z}) - \mathcal{E}(\mathbf{x}|\mathbf{z})] - \hat{\mathbf{x}}||_{\mathbb{X}}^2 \} \end{aligned} \quad (3.114)$$

$$= E_{\mathbf{x}, \mathbf{z}} \{ \langle \mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z}) + \mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}}, \mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z}) + \mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}} \rangle_{\mathbb{X}} \} \quad (3.115)$$

$$\begin{aligned} = E_{\mathbf{x}, \mathbf{z}} \{ \langle \mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z}), \mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z}) \rangle_{\mathbb{X}} + \langle \mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z}), \mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}} \rangle_{\mathbb{X}} \\ + \langle \mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}}, \mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z}) \rangle_{\mathbb{X}} + \langle \mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}}, \mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}} \rangle_{\mathbb{X}} \} \end{aligned} \quad (3.116)$$

Then, taking the expectation over each term yields

$$\begin{aligned} E_{\mathbf{x}, \mathbf{z}} (||\mathbf{x} - \hat{\mathbf{x}}||_{\mathbb{X}}^2) \\ = E_{\mathbf{x}, \mathbf{z}} \{ \langle \mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z}), \mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z}) \rangle_{\mathbb{X}} \} + E_{\mathbf{x}, \mathbf{z}} \{ \langle \mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z}), \mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}} \rangle_{\mathbb{X}} \} \\ + E_{\mathbf{x}, \mathbf{z}} \{ \langle \mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}}, \mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z}) \rangle_{\mathbb{X}} \} + E_{\mathbf{x}, \mathbf{z}} \{ \langle \mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}}, \mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}} \rangle_{\mathbb{X}} \} \end{aligned} \quad (3.117)$$

By Theorem 65, we note that $[\mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z})] \perp \mathcal{E}(\mathbf{x}|\mathbf{z})$ and $[\mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z})] \perp \hat{\mathbf{x}}$, thus the cross terms are zero. Note that $\hat{\mathbf{x}}$ is a statistic since it is a function of only the data. With the cross terms gone and using the induced norm notation for the remaining terms, we get

$$E_{\mathbf{x},\mathbf{z}}(\|\mathbf{x} - \hat{\mathbf{x}}\|_{\mathbb{X}}^2) = E_{\mathbf{x},\mathbf{z}}\{\|\mathbf{x} - \mathcal{E}(\mathbf{x}|\mathbf{z})\|_{\mathbb{X}}^2\} + E_{\mathbf{x},\mathbf{z}}\{\|\mathcal{E}(\mathbf{x}|\mathbf{z}) - \hat{\mathbf{x}}\|_{\mathbb{X}}^2\} \quad (3.118)$$

Thus, by inspection, $\hat{\mathbf{x}} = \mathcal{E}(\mathbf{x}|\mathbf{z})$ minimizes Equation (3.118) since the first term does not depend on $\hat{\mathbf{x}}$. ■

Therefore, this MMSE technique can be used to find a linear MVU estimator $\hat{\mathbf{x}} = g_o \circ \mathbf{z}$; when the random vectors are jointly Gaussian, then we can obtain the optimal MVU estimator without confining attention to only the optimum from the class of *linear* estimators. Since many of the problems of interest to us feature a generalized linear (or affine) relationship between the states and the observations, we will eventually pose our estimation problem in terms of a linear measurement model — see Definition 72. However, at this early stage of the development, we need only assume that our observations and states are statistically correlated with nonzero cross-covariances so that the observations contain information about the states; this is a necessary condition for using the conditional expectation to find a (linear) MVU estimator.

Definition 70 (Correlated States and Observations Measurement Model)

Let the random vector $\mathbf{z} \in \tilde{\mathbb{Z}} = \mathbb{L}^2(\Omega, P; \mathbb{Z})$ be the observation of a noise-corrupted measurement, $\mathbf{x} \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$ be the random state that is to be estimated, and \mathbb{Z} and \mathbb{X} be separable Hilbert spaces. Furthermore, we have knowledge of the means

$$\begin{aligned} \mu_{\mathbf{z}} &\triangleq E(\mathbf{z}) \\ \mu_{\mathbf{x}} &\triangleq E(\mathbf{x}) \end{aligned} \quad (3.119)$$

the nonzero covariances¹²

$$\begin{aligned}\Sigma(\mathbf{z}) &\triangleq E[(\mathbf{z} - \boldsymbol{\mu}_z) \diamond (\mathbf{z} - \boldsymbol{\mu}_z)] \\ \Sigma(\mathbf{x}) &\triangleq E[(\mathbf{x} - \boldsymbol{\mu}_x) \diamond (\mathbf{x} - \boldsymbol{\mu}_x)]\end{aligned}\tag{3.120}$$

and the nonzero cross-covariances

$$\begin{aligned}\Sigma(\mathbf{x}, \mathbf{z}) &\triangleq E[(\mathbf{x} - \boldsymbol{\mu}_x) \diamond (\mathbf{z} - \boldsymbol{\mu}_z)] \\ \Sigma(\mathbf{z}, \mathbf{x}) &\triangleq E[(\mathbf{z} - \boldsymbol{\mu}_z) \diamond (\mathbf{x} - \boldsymbol{\mu}_x)]\end{aligned}\tag{3.121}$$

The boxology for this measurement model is illustrated in Figure 3.6.

Finally, we come to our first result in this work that applies equally to problems with possibly nonzero-mean random vectors, a LIMVUE for CSO — this is the first step in our development of the ISKF.

Theorem 71 (LIMVUE for CSO) *Given the measurement model in Definition 70 and assuming the inverse $\Sigma^{-1}(\mathbf{z})$ exists¹³, then the state estimator, denoted by $\hat{\mathbf{x}} = \mathcal{E}(\mathbf{x}|\mathbf{z})$, the conditional expectation of the state, \mathbf{x} , given an observation, \mathbf{z} , is found by minimizing the MSE $E_{\mathbf{x},\mathbf{z}}(\|\mathbf{x} - \hat{\mathbf{x}}\|_{\mathbb{X}}^2)$ and is given by the general linear form [33]*

$$\hat{\mathbf{x}} = \mathbf{K}\mathbf{z} + \mathbf{c}\tag{3.122}$$

where

$$\mathbf{K} = \Sigma(\mathbf{x}, \mathbf{z}) \Sigma^{-1}(\mathbf{z})\tag{3.123}$$

¹²We have restricted ourselves to Hilbert spaces of Lebesgue \mathbb{L}^2 functions to guarantee the existence of covariances; however, this does not imply that their inverses exist.

¹³Even if the inverse does not exist, we can oftentimes find a suitable pseudoinverse [33] and still employ this estimator. Note that this estimator will no longer be “optimal” in any sense, but it may be a useful suboptimal algorithm nonetheless.

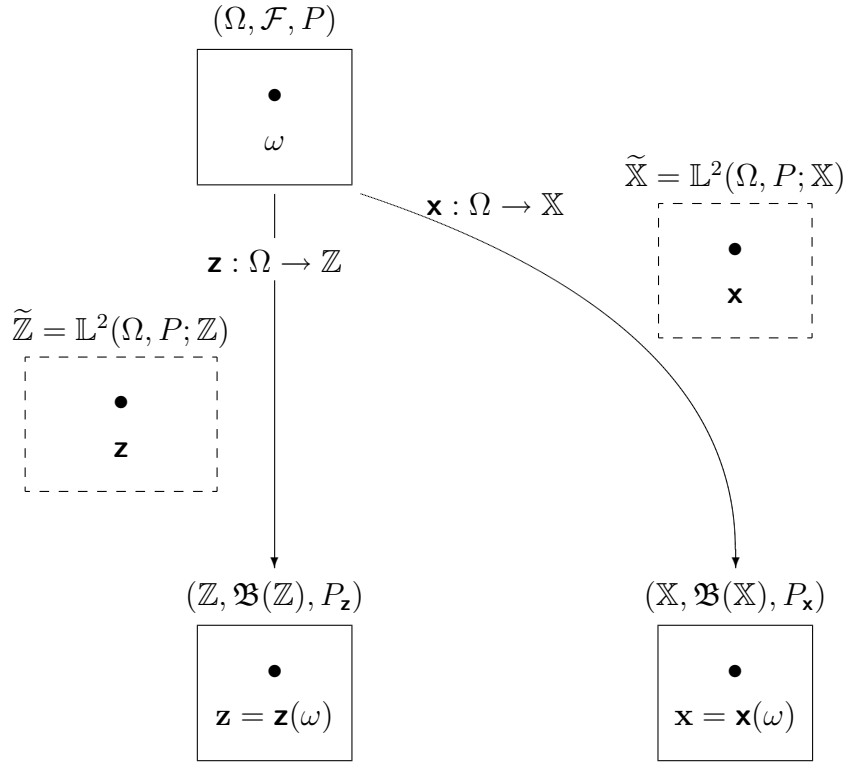


Figure 3.6 Boxology of a Measurement Model. The lack of symmetry in this figure is due to the process of “peeling” off information from the linear measurement model boxology shown in Figure 3.8 on page 3-57.

and the estimator bias term is

$$\mathbf{c} = \boldsymbol{\mu}_{\mathbf{x}} - \mathbf{K}\boldsymbol{\mu}_{\mathbf{z}} \quad (3.124)$$

The error, $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$, is zero-mean and has covariance

$$\boldsymbol{\Sigma}(\mathbf{e}) = \boldsymbol{\Sigma}(\mathbf{x}) - \mathbf{K}\boldsymbol{\Sigma}(\mathbf{z}, \mathbf{x}) \quad (3.125)$$

Remarks (1) We have not made any assumptions on the dimension of the state or observation. Thus $\boldsymbol{\Sigma}(\mathbf{x})$, $\boldsymbol{\Sigma}(\mathbf{z})$, and $\boldsymbol{\Sigma}(\mathbf{e})$ are covariance operators, while $\boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z})$ and $\boldsymbol{\Sigma}(\mathbf{z}, \mathbf{x})$ are cross-covariance transformations as described in Definition 45, page 3-28. (2) The boxology for this estimator appears in Figure 3.7. A close look tells

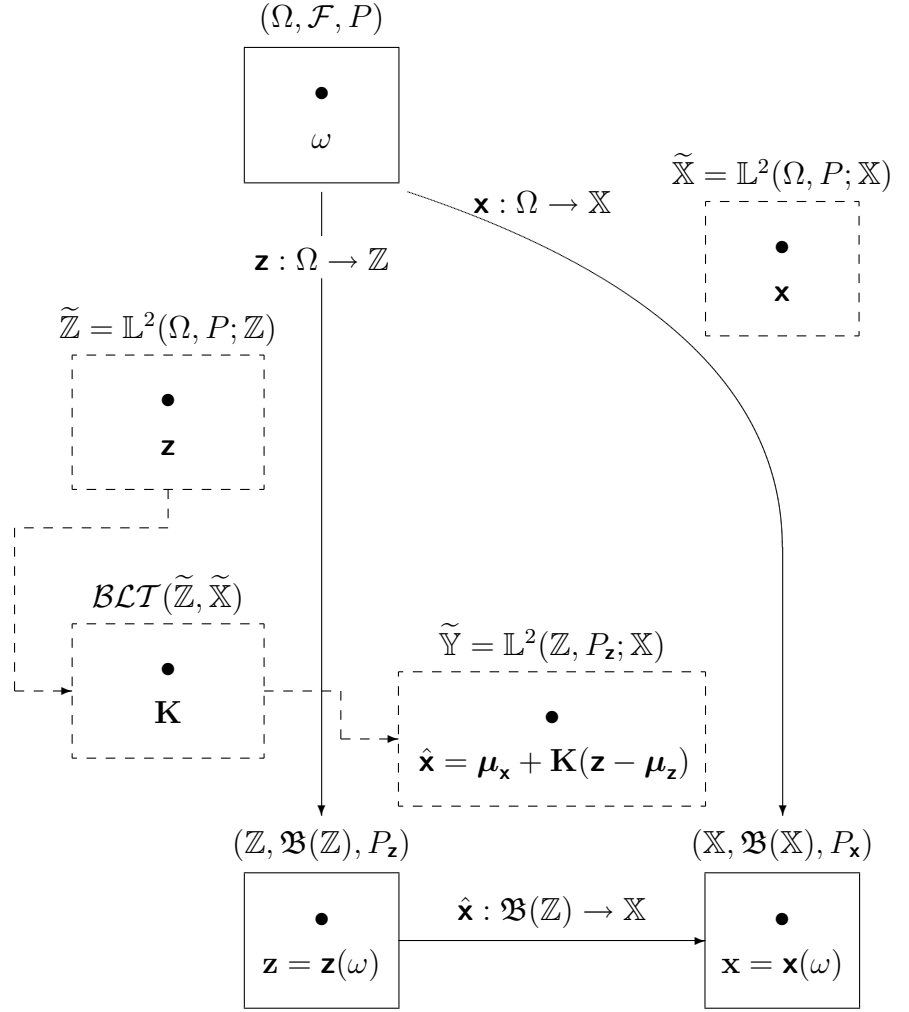


Figure 3.7 Boxology of the LIMVUE for CSO. The lack of symmetry in this figure is due to the process of “peeling” off information from the LIMVUE boxology shown in Figure 3.9 on page 3-63.

us quite clearly that we (a) have not assumed a *form* (linear or otherwise) for the relationship between the state and the observations, but (b) have assumed that the estimator is affine.

Proof of Theorem 71 First, we will find the estimator bias term, \mathbf{c} . Next, we will show that the linear state estimator, $\hat{\mathbf{x}}$, given by Equation (3.122), is unbiased. Then we shall derive the optimal transformation \mathbf{K} . Finally, we will show that the error covariance, given in Equation (3.125), is correct.

To find \mathbf{c} , we will minimize¹⁴ $\mathfrak{J}(\mathbf{c}) = E \left(\|\mathbf{x} - \hat{\mathbf{x}}\|^2 \right) = E \left[\|\mathbf{x} - (\mathbf{K}\mathbf{z} + \mathbf{c})\|^2 \right]$ with respect to \mathbf{c} . We will use the same technique as we used to show that the conditional mean is the MMSE estimator in Lemma 69. First, let $\mathbf{y} = \mathbf{x} - \mathbf{K}\mathbf{z}$, thus

$$\mathfrak{J}(\mathbf{c}) = E \left[\|\mathbf{x} - (\mathbf{K}\mathbf{z} + \mathbf{c})\|_{\mathbb{X}}^2 \right] = E \left[\|\mathbf{y} - \mathbf{c}\|_{\mathbb{X}}^2 \right] \quad (3.126)$$

Now expand the right-hand side with a wisely chosen zero $\boldsymbol{\mu}_{\mathbf{y}} - \boldsymbol{\mu}_{\mathbf{y}}$, where

$$\boldsymbol{\mu}_{\mathbf{y}} = E(\mathbf{y}) = E(\mathbf{x} - \mathbf{K}\mathbf{z}) = \boldsymbol{\mu}_{\mathbf{x}} - \mathbf{K}\boldsymbol{\mu}_{\mathbf{z}} \quad (3.127)$$

to obtain

$$\mathfrak{J}(\mathbf{c}) = E \left(\|\mathbf{y} + \boldsymbol{\mu}_{\mathbf{y}} - \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c}\|_{\mathbb{X}}^2 \right) \quad (3.128)$$

$$= E \left(\langle \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}} + \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c}, \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}} + \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c} \rangle_{\mathbb{X}} \right) \quad (3.129)$$

where the second line follows from the definition of the induced norm — on an inner product space, the norm squared is equal to the inner product of the quantity with itself. Continuing to expand we get

$$\begin{aligned} \mathfrak{J}(\mathbf{c}) &= E \left(\langle \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}, \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}} \rangle_{\mathbb{X}} + \langle \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}, \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c} \rangle_{\mathbb{X}} \right. \\ &\quad \left. + \langle \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c}, \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}} \rangle_{\mathbb{X}} + \langle \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c}, \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c} \rangle_{\mathbb{X}} \right) \end{aligned} \quad (3.130)$$

$$\begin{aligned} &= E \left(\langle \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}, \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}} \rangle_{\mathbb{X}} \right) + E \left(\langle \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}, \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c} \rangle_{\mathbb{X}} \right) \\ &\quad + E \left(\langle \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c}, \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}} \rangle_{\mathbb{X}} \right) + E \left(\langle \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c}, \boldsymbol{\mu}_{\mathbf{y}} - \mathbf{c} \rangle_{\mathbb{X}} \right) \end{aligned} \quad (3.131)$$

¹⁴Note that we have minimized the MSE using the joint expectation, $E_{\mathbf{x},\mathbf{z}}(\cdot)$.

Simplify Equation (3.131) by noting that the cross terms are equal to zero since $E(\mathbf{y} - \boldsymbol{\mu}_y) = \mathbf{0}$ and $\boldsymbol{\mu}_y - \mathbf{c}$ is a nonrandom quantity, therefore

$$\mathfrak{J}(\mathbf{c}) = E(\langle \mathbf{y} - \boldsymbol{\mu}_y, \mathbf{y} - \boldsymbol{\mu}_y \rangle_{\mathbb{X}}) + \langle \boldsymbol{\mu}_y - \mathbf{c}, \boldsymbol{\mu}_y - \mathbf{c} \rangle_{\mathbb{X}} \quad (3.132)$$

$$= E(\|\mathbf{y} - \boldsymbol{\mu}_y\|_{\mathbb{X}}^2) + \|\boldsymbol{\mu}_y - \mathbf{c}\|_{\mathbb{X}}^2 \quad (3.133)$$

where the first term, which is independent of \mathbf{c} , therefore $\mathfrak{J}(\mathbf{c})$ is minimized when $\|\boldsymbol{\mu}_y - \mathbf{c}\|_{\mathbb{X}}^2$ is equal to zero, which occurs when $\boldsymbol{\mu}_y - \mathbf{c} = \mathbf{0}$, hence $\mathbf{c} = \boldsymbol{\mu}_y = \boldsymbol{\mu}_x - \mathbf{K}\boldsymbol{\mu}_z$. Consequently, we can now write the state estimator, $\hat{\mathbf{x}} = \mathbf{K}\mathbf{z} + \mathbf{c}$, as

$$\hat{\mathbf{x}} = \boldsymbol{\mu}_x + \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_z) \quad (3.134)$$

The estimator bias is determined by taking the joint expectation of the error $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ [14]. Thus,

$$\text{bias}(\hat{\mathbf{x}}) = E(\mathbf{x} - \hat{\mathbf{x}}) = E\{\mathbf{x} - [\boldsymbol{\mu}_x + \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_z)]\} \quad (3.135)$$

Rearranging yields

$$\text{bias}(\hat{\mathbf{x}}) = E[(\mathbf{x} - \boldsymbol{\mu}_x) - \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_z)] \quad (3.136)$$

Finally, moving the expectation inside yields

$$\text{bias}(\hat{\mathbf{x}}) = E(\mathbf{x} - \boldsymbol{\mu}_x) - \mathbf{K}E(\mathbf{z} - \boldsymbol{\mu}_z) = \mathbf{0} \quad (3.137)$$

Therefore, the estimator is unbiased as claimed.

Now we seek the transformation $\mathbf{K} \in \mathcal{BLT}(\tilde{\mathbb{Z}}, \tilde{\mathbb{X}})$, as shown in Figure 3.7, that minimizes the expected value of the squared error between our affine estimator $\hat{\mathbf{x}} = \boldsymbol{\mu}_x + \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_z)$, given observation \mathbf{z} , and the state \mathbf{x} . Thus, the cost function that we want to minimize is $\mathfrak{J}(\mathbf{K}) = E\{\|\mathbf{x} - [\boldsymbol{\mu}_x + \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_z)]\|^2\}$; hence our task is to minimize $\mathfrak{J}(\mathbf{K})$ with respect to \mathbf{K} . We begin by writing and then expanding the

expression for $\mathfrak{J}(\mathbf{K})$

$$\mathfrak{J}(\mathbf{K}) = E[||(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) - \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})||^2] \quad (3.138)$$

$$= E[\langle (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) - \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}), (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) - \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \rangle] \quad (3.139)$$

$$= \text{tr } E\{[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) - \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})] \diamond [(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) - \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})]\} \quad (3.140)$$

where the second line follows from the definition of the induced norm and the third line is a powerful identity that relates an inner product to the trace of an outer product (for any type of expectation operator) as stated in Equation (3.75) on page 3-31. Note that $E\{[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) - \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})] \diamond [(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) - \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})]\}$ is the error correlation, $\boldsymbol{\Xi}(\mathbf{e})$, or equivalently, the error covariance, $\boldsymbol{\Sigma}(\mathbf{e})$, since the error is zero-mean. Continuing,

$$\begin{aligned} \mathfrak{J}(\mathbf{K}) &= \text{tr } E\{[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] - [(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})] \\ &\quad - [\mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + [\mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \diamond \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})]\} \end{aligned} \quad (3.141)$$

$$\begin{aligned} &= \text{tr}\{E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] - E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})] \\ &\quad - E[\mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + E[\mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \diamond \mathbf{K}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})]\} \end{aligned} \quad (3.142)$$

$$\begin{aligned} &= \text{tr}\{E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] - E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})]\mathbf{K}^* \\ &\quad - \mathbf{K}E[(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + \mathbf{K}E[(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \diamond (\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})]\mathbf{K}^*\} \end{aligned} \quad (3.143)$$

$$= \text{tr}[\boldsymbol{\Sigma}(\mathbf{x}) - \boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z})\mathbf{K}^* - \mathbf{K}\boldsymbol{\Sigma}(\mathbf{z}, \mathbf{x}) + \mathbf{K}\boldsymbol{\Sigma}(\mathbf{z})\mathbf{K}^*] \quad (3.144)$$

where the second line is due to the linearity of the expectation operator and the third line is due to Lemma 43, page 3-26; \mathbf{K}^* is the adjoint of transformation \mathbf{K} ; and covariance operator notation is used in the fourth line.

If \mathfrak{J} were a function of a scalar, vector, or matrix, then we could use standard calculus to minimize the function; however, in this problem, \mathfrak{J} is function of a transformation and hence, the rules of calculus [5] and/or vector calculus [126] do not apply. So, we will use a calculus of variations method [208, 60, 122, 8] to minimize a

function with respect to a transformation. For positive $\alpha \in \mathbb{R}$ and transformations $\mathbf{K}, \mathbf{L} \in \mathcal{BCT}(\tilde{\mathbb{Z}}, \tilde{\mathbb{X}})$, the Gateaux variation, $\delta\mathfrak{J}(\mathbf{K}; \mathbf{L})$, defined as

$$\delta\mathfrak{J}(\mathbf{K}; \mathbf{L}) = \lim_{\alpha \rightarrow 0} \frac{\mathfrak{J}(\mathbf{K} + \alpha\mathbf{L}) - \mathfrak{J}(\mathbf{K})}{\alpha} \quad (3.145)$$

can be used to minimize $\mathfrak{J}(\mathbf{K})$ with respect to \mathbf{K} . To find the optimal \mathbf{K} , we assume that \mathbf{K} is optimal when the variation is zero for all \mathbf{L} — then we solve for the optimal \mathbf{K} — this is a necessary optimality condition [208, 122]. Substituting $\mathbf{K} + \alpha\mathbf{L}$ in for \mathbf{K} in Equation (3.144) we can write

$$\begin{aligned} \mathfrak{J}(\mathbf{K} + \alpha\mathbf{L}) &= \text{tr}[\Sigma(\mathbf{x}) - \Sigma(\mathbf{x}, \mathbf{z}) (\mathbf{K} + \alpha\mathbf{L})^* \\ &\quad - (\mathbf{K} + \alpha\mathbf{L}) \Sigma(\mathbf{z}, \mathbf{x}) + (\mathbf{K} + \alpha\mathbf{L}) \Sigma(\mathbf{z}) (\mathbf{K} + \alpha\mathbf{L})^*] \end{aligned} \quad (3.146)$$

$$\begin{aligned} &= \text{tr}[\Sigma(\mathbf{x}) - \Sigma(\mathbf{x}, \mathbf{z}) \mathbf{K}^* - \Sigma(\mathbf{x}, \mathbf{z}) \alpha\mathbf{L}^* \\ &\quad - \mathbf{K} \Sigma(\mathbf{z}, \mathbf{x}) - \alpha\mathbf{L} \Sigma(\mathbf{z}, \mathbf{x}) + \mathbf{K} \Sigma(\mathbf{z}) \mathbf{K}^* \\ &\quad + \mathbf{K} \Sigma(\mathbf{z}) \alpha\mathbf{L}^* + \alpha\mathbf{L} \Sigma(\mathbf{z}) \mathbf{K}^* + \alpha\mathbf{L} \Sigma(\mathbf{z}) \alpha\mathbf{L}^*] \end{aligned} \quad (3.147)$$

Subtracting Equation (3.144) from Equation (3.147) yields

$$\begin{aligned} \mathfrak{J}(\mathbf{K} + \alpha\mathbf{L}) - \mathfrak{J}(\mathbf{K}) &= \text{tr}[\mathbf{K} \Sigma(\mathbf{z}) \alpha\mathbf{L}^* + \alpha\mathbf{L} \Sigma(\mathbf{z}) \mathbf{K}^* + \alpha\mathbf{L} \Sigma(\mathbf{z}) \alpha\mathbf{L}^* \\ &\quad - \alpha\mathbf{L} \Sigma(\mathbf{z}, \mathbf{x}) - \Sigma(\mathbf{x}, \mathbf{z}) \alpha\mathbf{L}^*] \end{aligned} \quad (3.148)$$

Divide by α to yield

$$\begin{aligned} \frac{\mathfrak{J}(\mathbf{K} + \alpha\mathbf{L}) - \mathfrak{J}(\mathbf{K})}{\alpha} &= \text{tr}[\mathbf{K} \Sigma(\mathbf{z}) \mathbf{L}^* + \mathbf{L} \Sigma(\mathbf{z}) \mathbf{K}^* + \mathbf{L} \Sigma(\mathbf{z}) \mathbf{L}^* \\ &\quad - \mathbf{L} \Sigma(\mathbf{z}, \mathbf{x}) - \Sigma(\mathbf{x}, \mathbf{z}) \mathbf{L}^*] \end{aligned} \quad (3.149)$$

and take the limit as $\alpha \rightarrow 0$: then Equation (3.145) becomes

$$\delta\mathfrak{J}(\mathbf{K}; \mathbf{L}) = \text{tr}[\mathbf{K} \Sigma(\mathbf{z}) \mathbf{L}^* + \mathbf{L} \Sigma(\mathbf{z}) \mathbf{K}^* - \mathbf{L} \Sigma(\mathbf{z}, \mathbf{x}) - \Sigma(\mathbf{x}, \mathbf{z}) \mathbf{L}^*] \quad (3.150)$$

Factoring out \mathbf{L}^* and \mathbf{L} results in

$$\delta\mathfrak{J}(\mathbf{K}; \mathbf{L}) = \text{tr}\{[\mathbf{K} \boldsymbol{\Sigma}(\mathbf{z}) - \boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z})]\mathbf{L}^* + \mathbf{L}[\boldsymbol{\Sigma}(\mathbf{z}) \mathbf{K}^* - \boldsymbol{\Sigma}(\mathbf{z}, \mathbf{x})]\} \quad (3.151)$$

and since $\text{tr}\mathbf{A}^* = \text{tr}\mathbf{A}$,

$$\delta\mathfrak{J}(\mathbf{K}; \mathbf{L}) = 2\text{tr}\{[\mathbf{K} \boldsymbol{\Sigma}(\mathbf{z}) - \boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z})]\mathbf{L}^*\} \quad (3.152)$$

Assume the optimal \mathbf{K}_o minimizes \mathfrak{J} . Then a necessary optimality condition [122] yields the differential $\delta\mathfrak{J}(\mathbf{K}_o; \mathbf{L}) = 0$ for all \mathbf{L} , and we get

$$\text{tr}\{[\mathbf{K}_o \boldsymbol{\Sigma}(\mathbf{z}) - \boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z})]\mathbf{L}^*\} = 0 \text{ for all } \mathbf{L} \quad (3.153)$$

which implies that

$$\mathbf{K}_o \boldsymbol{\Sigma}(\mathbf{z}) - \boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z}) = \mathbf{0} \quad (3.154)$$

Rearranging and assuming that our measurement covariance, $\boldsymbol{\Sigma}(\mathbf{z})$, is invertible¹⁵

$$\mathbf{K}_o = \boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z})\boldsymbol{\Sigma}^{-1}(\mathbf{z}) \quad (3.155)$$

Hence the state estimator, $\hat{\mathbf{x}} = \boldsymbol{\mu}_x + \mathbf{K}_o(\mathbf{z} - \boldsymbol{\mu}_z)$, from Equation (3.134), becomes

$$\hat{\mathbf{x}} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z}) \boldsymbol{\Sigma}^{-1}(\mathbf{z}) (\mathbf{z} - \boldsymbol{\mu}_z) \quad (3.156)$$

¹⁵Per Definition 45, the covariance operator is only guaranteed to be positive, which means that it may have one (or more) eigenvalues which are zero. To be invertible, an operator may not have any zero eigenvalues. However, in a practical system, the measurements are not perfect and hence all of the eigenvalues are positive, versus nonnegative. Therefore, it not restrictive to assume that $\boldsymbol{\Sigma}(\mathbf{z})$ is invertible. Additionally, we later show that, for the case of a linear measurement model with additive noise, it is sufficient to assume that the noise covariance operator is strictly positive to guarantee the invertibility of $\boldsymbol{\Sigma}(\mathbf{z})$.

Since Equation (3.144) is the trace of the error covariance, we have

$$\Sigma(\mathbf{e}) = \Sigma(\mathbf{x}) - \Sigma(\mathbf{x}, \mathbf{z}) \mathbf{K}^* - \mathbf{K} \Sigma(\mathbf{z}, \mathbf{x}) + \mathbf{K} \Sigma(\mathbf{z}) \mathbf{K}^* \quad (3.157)$$

$$= \Sigma(\mathbf{x}) - \mathbf{K} \Sigma(\mathbf{z}, \mathbf{x}) + [\mathbf{K} \Sigma(\mathbf{z}) - \Sigma(\mathbf{x}, \mathbf{z})] \mathbf{K}^* \quad (3.158)$$

Using the expression in Equation (3.154) yields

$$\Sigma(\mathbf{e}) = \Sigma(\mathbf{x}) - \mathbf{K} \Sigma(\mathbf{z}, \mathbf{x}) \quad (3.125)$$

Therefore, all of the parts of this theorem have been proved. ■

Next, we employ a generalized linear measurement model with zero-mean white Gaussian additive noise (WGAN).

Definition 72 (Generalized Linear Measurement Model) *The generalized linear measurement model is represented by the algebraic equation*

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v} \quad (3.159)$$

where:

$$\begin{aligned} \mathbf{z} &\in \tilde{\mathbb{Z}} = \mathbb{L}^2(\Omega, P; \mathbb{Z}) && \dots \text{ measurement vector} \\ \mathbf{H} &\in \mathcal{BLT}(\tilde{\mathbb{X}}, \tilde{\mathbb{Z}}) && \dots \text{ measurement distributor transformation} \\ \mathbf{x} &\in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X}) && \dots \text{ state vector} \\ \mathbf{v} &\in \tilde{\mathbb{V}} = \mathbb{L}^2(\Omega, P; \mathbb{V}) && \dots \text{ measurement-corruption noise vector} \end{aligned}$$

and \mathbf{H} is known and \mathbb{Z} , \mathbb{X} , and \mathbb{V} are separable Hilbert spaces. Additionally, the measurement is corrupted by zero-mean white Gaussian additive noise, \mathbf{v} , with a known covariance operator

$$\Sigma(\mathbf{v}) = E(\mathbf{v} \diamond \mathbf{v}) = \mathbf{R} \in \mathcal{BLO}(\tilde{\mathbb{V}}) \quad (3.160)$$

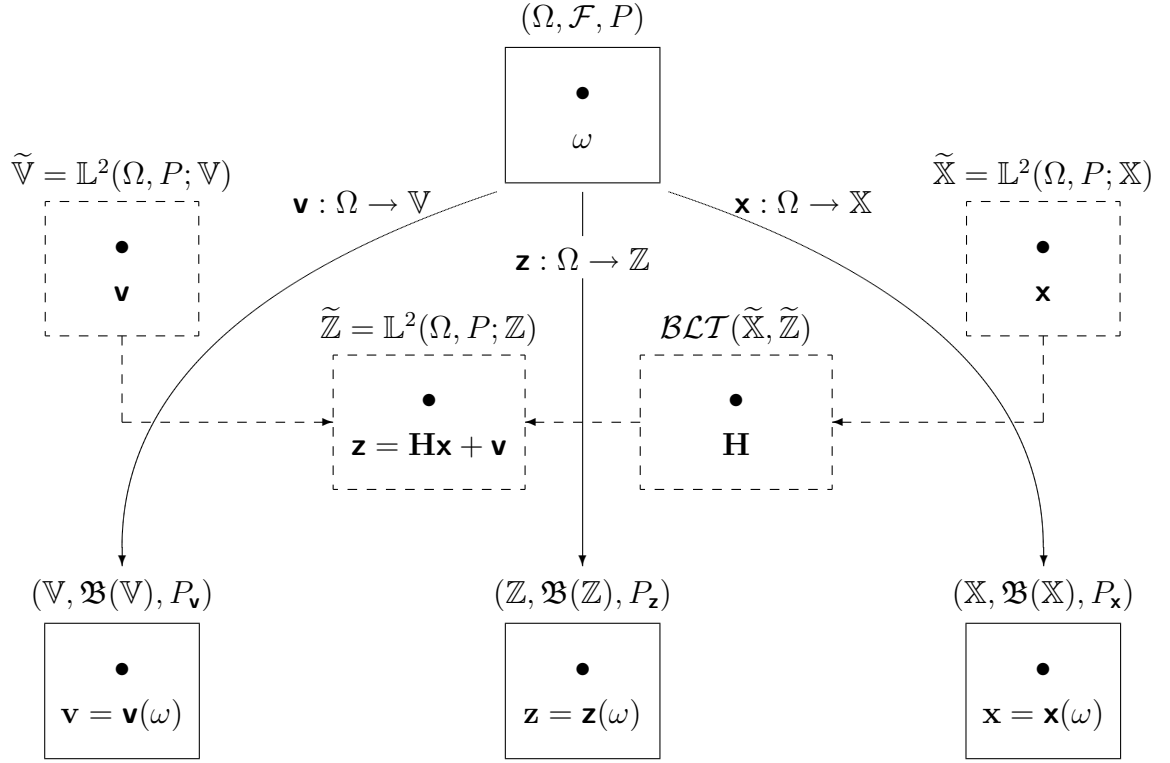


Figure 3.8 Boxology of a Generalized Linear Measurement Model.

which is symmetric and nuclear since it is a covariance operator (see Definition 45) and strictly positive by the nonrestrictive reasoning given in the second remark following Definition 62. The mean of the state is $\mu_{\mathbf{x}}$ and the covariance operator for the random state vector given by

$$\Sigma(\mathbf{x}) = E[(\mathbf{x} - \mu_{\mathbf{x}}) \diamond (\mathbf{x} - \mu_{\mathbf{x}})] = \mathbf{P} \in \mathcal{BLO}(\tilde{\mathbb{X}}) \quad (3.161)$$

is symmetric, positive, and nuclear per Definition 45. Furthermore, we note that the state and measurement-corruption noise are independent¹⁶.

The boxology for this linear measurement model is illustrated in Figure 3.8.

¹⁶The correctness of this statement will be demonstrated after we have completed the ISKF proof. It will be seen that the state and the measurement-corruption noise are independent because the measurement-corruption noise is mutually independent of the dynamics driving noise and the initial state.

The following lemma establishes the corresponding result for infinite-dimensional systems on Hilbert spaces, that independent random vectors defined on finite-dimensional Hilbert spaces are also uncorrelated, hence their cross-covariances are zero.

Lemma 73 *Let \mathbb{X} and \mathbb{V} be separable Hilbert spaces and $\mathbf{x} \in \mathfrak{F}(\Omega, \mathbb{X})$ and $\mathbf{v} \in \mathfrak{F}(\Omega, \mathbb{V})$ be \mathbb{X} and \mathbb{V} -valued random vectors, respectively. If \mathbf{x} and \mathbf{v} are independent, then they are also uncorrelated and thus the cross-covariance transformations are zero, i.e., $\Sigma(\mathbf{x}, \mathbf{v}) = \mathbf{0}$ and $\Sigma(\mathbf{v}, \mathbf{x}) = \mathbf{0}$.*

Proof of Lemma 73 From Definition 45, the cross-covariance of random vectors \mathbf{x} and \mathbf{v} is given by

$$\Sigma(\mathbf{x}, \mathbf{v}) \triangleq E[(\mathbf{x} - \mu_{\mathbf{x}}) \diamond (\mathbf{v} - \mu_{\mathbf{v}})] \quad (3.162)$$

$$= E[(\mathbf{x} \diamond \mathbf{v}) - (\mu_{\mathbf{x}} \diamond \mathbf{v}) - (\mathbf{x} \diamond \mu_{\mathbf{v}}) + (\mu_{\mathbf{x}} \diamond \mu_{\mathbf{v}})] \quad (3.163)$$

$$= E(\mathbf{x} \diamond \mathbf{v}) - E(\mu_{\mathbf{x}} \diamond \mathbf{v}) - E(\mathbf{x} \diamond \mu_{\mathbf{v}}) + E(\mu_{\mathbf{x}} \diamond \mu_{\mathbf{v}}) \quad (3.164)$$

We shall show, in turn, that all four terms of Equation (3.164) are equal to $\mu_{\mathbf{x}} \diamond \mu_{\mathbf{v}}$. The first term, $E(\mathbf{x} \diamond \mathbf{v})$, is the correlation transformation of random vector \mathbf{x} with random noise vector \mathbf{v} . For all $\boldsymbol{\eta} \in \mathbb{V}$ we have

$$[E(\mathbf{x} \diamond \mathbf{v})]\boldsymbol{\eta} = E[(\mathbf{x} \diamond \mathbf{v})\boldsymbol{\eta}] = E[\mathbf{x}\langle\mathbf{v}, \boldsymbol{\eta}\rangle] \quad (3.165)$$

where the last equality follows from the definition of the outer product. Since \mathbf{x} and \mathbf{v} are independent, then \mathbf{x} and a linear function of \mathbf{v} , namely the functional $\langle\mathbf{v}, \boldsymbol{\eta}\rangle$, are independent, thus

$$[E(\mathbf{x} \diamond \mathbf{v})]\boldsymbol{\eta} = E(\mathbf{x})E(\langle\mathbf{v}, \boldsymbol{\eta}\rangle) = \mu_{\mathbf{x}}\langle E(\mathbf{v}), \boldsymbol{\eta}\rangle = (\mu_{\mathbf{x}} \diamond \mu_{\mathbf{v}})\boldsymbol{\eta} \quad (3.166)$$

for all $\boldsymbol{\eta} \in \mathbb{V}$. Therefore, we get $E(\mathbf{x} \diamond \mathbf{v}) = \mu_{\mathbf{x}} \diamond \mu_{\mathbf{v}}$.

The second and third terms on the right-hand side of Equation (3.164) follow from similar arguments and, thus, only second term will be explicitly shown. For every $\boldsymbol{\eta} \in \mathbb{V}$,

$$[E(\boldsymbol{\mu}_{\mathbf{x}} \diamond \mathbf{v})]\boldsymbol{\eta} = E[(\boldsymbol{\mu}_{\mathbf{x}} \diamond \mathbf{v})\boldsymbol{\eta}] = E[\boldsymbol{\mu}_{\mathbf{x}}\langle \mathbf{v}, \boldsymbol{\eta} \rangle] \quad (3.167)$$

Then, since $\boldsymbol{\mu}_{\mathbf{x}}$ is not random, we obtain

$$[E(\boldsymbol{\mu}_{\mathbf{x}} \diamond \mathbf{v})]\boldsymbol{\eta} = \boldsymbol{\mu}_{\mathbf{x}}E(\langle \mathbf{v}, \boldsymbol{\eta} \rangle) = \boldsymbol{\mu}_{\mathbf{x}}\langle E(\mathbf{v}), \boldsymbol{\eta} \rangle = (\boldsymbol{\mu}_{\mathbf{x}} \diamond \boldsymbol{\mu}_{\mathbf{v}})\boldsymbol{\eta} \quad (3.168)$$

Thus, $E(\boldsymbol{\mu}_{\mathbf{x}} \diamond \mathbf{v}) = \boldsymbol{\mu}_{\mathbf{x}} \diamond \boldsymbol{\mu}_{\mathbf{v}}$ and $E(\mathbf{x} \diamond \boldsymbol{\mu}_{\mathbf{v}}) = \boldsymbol{\mu}_{\mathbf{x}} \diamond \boldsymbol{\mu}_{\mathbf{v}}$.

Clearly, $E(\boldsymbol{\mu}_{\mathbf{x}} \diamond \boldsymbol{\mu}_{\mathbf{v}}) = \boldsymbol{\mu}_{\mathbf{x}} \diamond \boldsymbol{\mu}_{\mathbf{v}}$, since the means are not random. Therefore, all four terms are $\boldsymbol{\mu}_{\mathbf{x}} \diamond \boldsymbol{\mu}_{\mathbf{v}}$ and the lemma holds. An analogous set of steps will lead us to the conclusion that $\boldsymbol{\Sigma}(\mathbf{v}, \mathbf{x}) = \mathbf{0}$. ■

In Theorem 75, we require that the measurement covariance operator, $\boldsymbol{\Sigma}(\mathbf{z}) = \mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}$, be invertible. Hence we shall attend to it now in the following lemma.

Lemma 74 (Measurement Covariance Operator) *Let \mathbf{H} , \mathbf{P} , and \mathbf{R} be as described in Definition 72, then the inverse of $\boldsymbol{\Sigma}(\mathbf{z}) = \mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}$ exists.*

Proof of Lemma 74 First, we show that $\boldsymbol{\Sigma}(\mathbf{z})$ is $\mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}$. We begin with the definition of the covariance operator and then substitute $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$ and

$$\boldsymbol{\mu}_{\mathbf{z}} = E(\mathbf{z}) = E(\mathbf{H}\mathbf{x} + \mathbf{v}) = \mathbf{H}E(\mathbf{x}) + E(\mathbf{v}) = \mathbf{H}\boldsymbol{\mu}_{\mathbf{x}} \quad (3.169)$$

into $\boldsymbol{\Sigma}(\mathbf{z})$ and then regroup terms to get

$$\boldsymbol{\Sigma}(\mathbf{z}) \triangleq E[(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \diamond (\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})] \quad (3.170)$$

$$= E[(\mathbf{H}\mathbf{x} + \mathbf{v} - \mathbf{H}\boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{H}\mathbf{x} + \mathbf{v} - \mathbf{H}\boldsymbol{\mu}_{\mathbf{x}})] \quad (3.171)$$

$$= E\{[\mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) + \mathbf{v}] \diamond [\mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) + \mathbf{v}]\} \quad (3.172)$$

Now employ the distributive property of the outer product (shown in Lemma 16, page 3-11) to obtain

$$\begin{aligned}\Sigma(\mathbf{z}) = & E\{[\mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond \mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + [\mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond \mathbf{v}] \\ & + [\mathbf{v} \diamond \mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + (\mathbf{v} \diamond \mathbf{v})\}\end{aligned}\quad (3.173)$$

Moving the expectation in, then using Lemma 43 on page 3-26 to factor the nonrandom operators out of the expectation yields

$$\begin{aligned}\Sigma(\mathbf{z}) = & E[\mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond \mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + E[\mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond \mathbf{v}] \\ & + E[\mathbf{v} \diamond \mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + E(\mathbf{v} \diamond \mathbf{v})\end{aligned}\quad (3.174)$$

$$\begin{aligned}= & \mathbf{H}E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})]\mathbf{H}^* + \mathbf{H}E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond \mathbf{v}] \\ & + E[\mathbf{v} \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})]\mathbf{H}^* + E(\mathbf{v} \diamond \mathbf{v})\end{aligned}\quad (3.175)$$

$$= \mathbf{H}\Sigma(\mathbf{x})\mathbf{H}^* + \mathbf{H}\Sigma(\mathbf{x}, \mathbf{v}) + \Sigma(\mathbf{v}, \mathbf{x})\mathbf{H}^* + \Sigma(\mathbf{v})\quad (3.176)$$

and the third line results from the definition of the covariance operator and the distributive property of the outer product.

Therefore, we can write Equation (3.176) as

$$\Sigma(\mathbf{z}) = \mathbf{H}\Sigma(\mathbf{x})\mathbf{H}^* + \Sigma(\mathbf{v}) = \mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}\quad (3.177)$$

since both cross-covariance terms are zero per Lemma 73.

Next we show that $\mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}$ has an inverse. In general, we know that a covariance operator is positive. Thus, a sufficient condition for our lemma is that either one of the terms must be strictly positive. By our assumed specifications on the measurement-corruption noise covariance, \mathbf{R} is strictly positive (or positive definite if \mathbf{R} is the matrix representation of an operator). Let's begin our proof by contradiction by stating that $\mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}$ is nonpositive, i.e., for every $\boldsymbol{\zeta} \in \tilde{\mathbb{Z}}$ the

following holds

$$\langle (\mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R})\boldsymbol{\zeta}, \boldsymbol{\zeta} \rangle_{\tilde{\mathbb{Z}}} \leq 0 \quad (3.178)$$

Using additivity property of the inner product we get

$$\langle \mathbf{H}\mathbf{P}\mathbf{H}^*\boldsymbol{\zeta}, \boldsymbol{\zeta} \rangle_{\tilde{\mathbb{Z}}} + \langle \mathbf{R}\boldsymbol{\zeta}, \boldsymbol{\zeta} \rangle_{\tilde{\mathbb{Z}}} \leq 0 \quad (3.179)$$

We know that the second term, $\langle \mathbf{R}\boldsymbol{\zeta}, \boldsymbol{\zeta} \rangle_{\tilde{\mathbb{Z}}}$, on the left-hand side of Equation (3.179) is greater than zero, hence the first term must be negative (a necessary condition) for the equation to hold true. For $\mathbf{H} \in \mathcal{BLT}(\tilde{\mathbb{X}}, \tilde{\mathbb{Z}})$, the adjoint of \mathbf{H} is $\mathbf{H}^* \in \mathcal{BLT}(\tilde{\mathbb{Z}}, \tilde{\mathbb{X}})$ and thus

$$\langle \mathbf{H}\mathbf{P}\mathbf{H}^*\boldsymbol{\zeta}, \boldsymbol{\zeta} \rangle_{\tilde{\mathbb{Z}}} = \langle \mathbf{P}\mathbf{H}^*\boldsymbol{\zeta}, \mathbf{H}^*\boldsymbol{\zeta} \rangle_{\tilde{\mathbb{X}}} \quad (3.180)$$

where $\mathbf{H}^*\boldsymbol{\zeta} = \boldsymbol{\xi}$ and $\boldsymbol{\xi} \in \tilde{\mathbb{X}}$, so Equation (3.180) becomes

$$\langle \mathbf{H}\mathbf{P}\mathbf{H}^*\boldsymbol{\zeta}, \boldsymbol{\zeta} \rangle_{\tilde{\mathbb{Z}}} = \langle \mathbf{P}\boldsymbol{\xi}, \boldsymbol{\xi} \rangle_{\tilde{\mathbb{X}}} \quad (3.181)$$

We know that the state covariance \mathbf{P} , defined in Equation (3.161) is positive, i.e., for every $\boldsymbol{\xi} \in \tilde{\mathbb{X}}$,

$$\langle \mathbf{P}\boldsymbol{\xi}, \boldsymbol{\xi} \rangle_{\tilde{\mathbb{X}}} \geq 0 \quad (3.182)$$

Thus, we have a contradiction and hence $\mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}$ is strictly positive and therefore invertible. ■

Now we are ready to solve a more specific problem using Theorem 71 in conjunction with the new generalized linear measurement model given in Definition 72. With this new measurement model, the LIMVUE for CSO now becomes the LIMVUE¹⁷ for the generalized linear measurement model.

¹⁷For finite-dimensional systems, the linear MVU estimator (LMVUE) is also the *best* linear unbiased estimator (BLUE), in the MMSE sense, for the class of linear and unbiased estimators. The LMVUE (or BLUE) with jointly Gaussian random vectors is the overall best (optimal) MVU estimator [100, 14]; hence there are no nonlinear estimators that are better, in the MMSE sense. We expect that *this* LIMVUE is the overall best infinite-dimensional MVU estimator (IMVUE).

Theorem 75 (LIMVUE) *We employ the generalized linear measurement model given in Definition 72 to describe the random vectors representing the measurement, state, and measurement-corruption noise. The LIMVUE is*

$$\hat{\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{x}} + \mathbf{P}\mathbf{H}^*[\mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}]^{-1}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \quad (3.183)$$

where $\boldsymbol{\mu}_{\mathbf{z}} = \mathbf{H}\boldsymbol{\mu}_{\mathbf{x}}$ and the error covariance operator is

$$\boldsymbol{\Sigma}(\mathbf{e}) = \mathbf{P} - \mathbf{P}\mathbf{H}^*[\mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}]^{-1}\mathbf{H}\mathbf{P} \quad (3.184)$$

Since the LIMVUE is simply the LIMVUE for CSO with the generalized linear measurement model, the boxology for the LIMVUE (illustrated in Figure 3.9) is a combination of the LIMVUE for CSO boxology, see Figure 3.7 on page 3-50, and the boxology of the generalized linear measurement model shown in Figure 3.8, page 3-57.

Proof of Theorem 75 Per Theorem 71, the estimator is indeed unbiased. We shall use Theorem 71 two more times to find the estimator and the error covariance for this theorem.

The LIMVUE for CSO, $\hat{\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z}) \boldsymbol{\Sigma}^{-1}(\mathbf{z}) (\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})$, as given in Equation (3.156), is our starting point. We will first determine a new expression for the cross-covariance $\boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z})$ by substituting $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$ and

$$\boldsymbol{\mu}_{\mathbf{z}} = E(\mathbf{z}) = E(\mathbf{H}\mathbf{x} + \mathbf{v}) = \mathbf{H}E(\mathbf{x}) + E(\mathbf{v}) = \mathbf{H}\boldsymbol{\mu}_{\mathbf{x}} \quad (3.185)$$

into the definition of $\boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z})$ to get

$$\boldsymbol{\Sigma}(\mathbf{x}, \mathbf{z}) \triangleq E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})] \quad (3.186)$$

$$= E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{H}\mathbf{x} + \mathbf{v} - \mathbf{H}\boldsymbol{\mu}_{\mathbf{x}})] \quad (3.187)$$

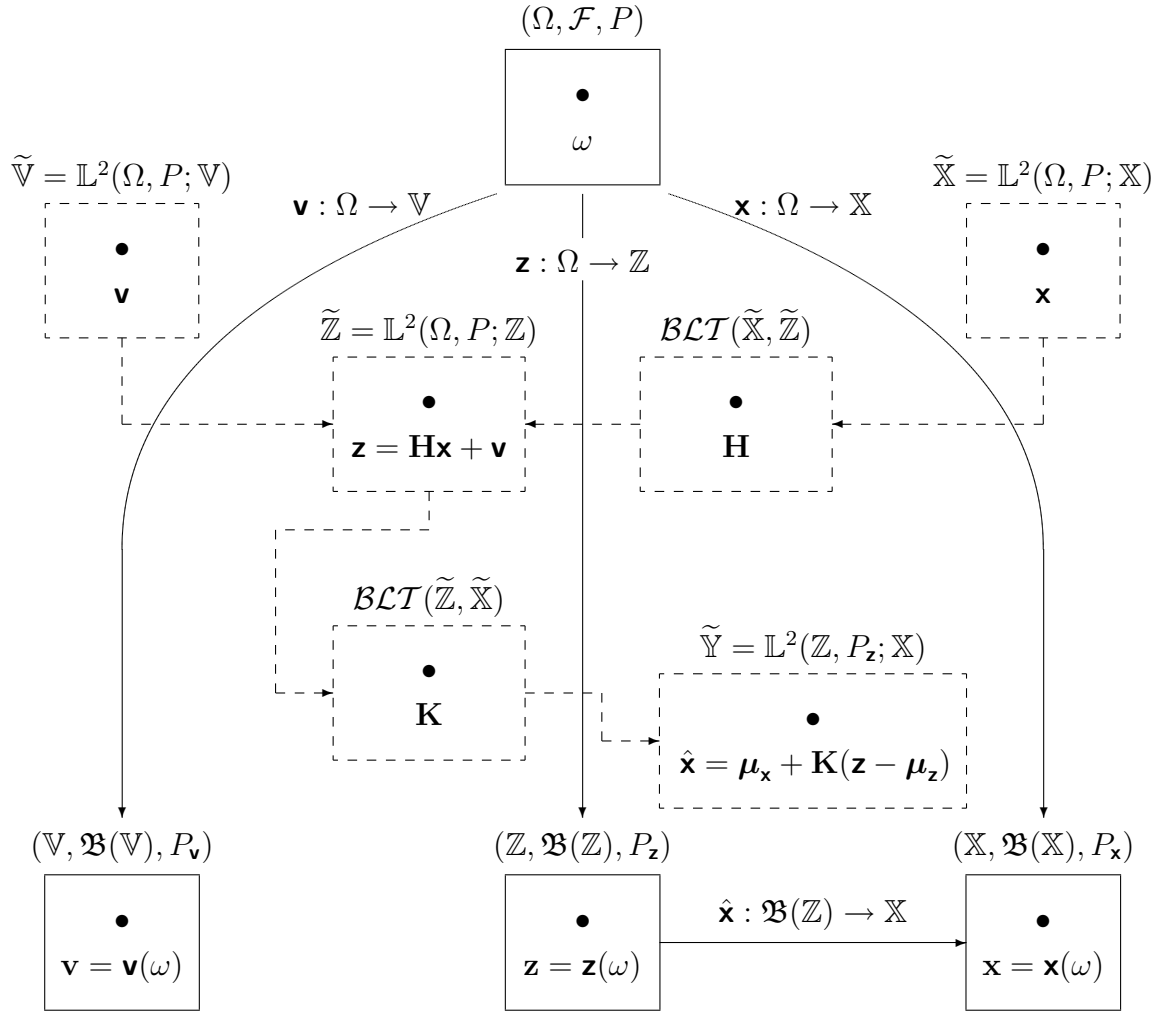


Figure 3.9 Boxology of the Linear Infinite-Dimensional Minimum Variance Unbiased Estimator.

Applying the distributive property in Lemma 16 (found on page 3-11) results in

$$\Sigma(\mathbf{x}, \mathbf{z}) = E\{[(\mathbf{x} - \mu_{\mathbf{x}}) \diamond (\mathbf{H}\mathbf{x} - \mathbf{H}\mu_{\mathbf{x}})] + [(\mathbf{x} - \mu_{\mathbf{x}}) \diamond \mathbf{v}]\} \quad (3.188)$$

Then it follows from the linearity of the expectation operator and, in the second line,

from the definition of the cross-covariance

$$\Sigma(\mathbf{x}, \mathbf{z}) = E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{H}\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_{\mathbf{x}})] + E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond \mathbf{v}] \quad (3.189)$$

$$= E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond \mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + \Sigma(\mathbf{x}, \mathbf{v}) \quad (3.190)$$

The cross-covariance $\Sigma(\mathbf{x}, \mathbf{v})$ is zero since \mathbf{x} and \mathbf{v} are independent (and thus uncorrelated)¹⁸ hence Equation (3.190) becomes

$$\Sigma(\mathbf{x}, \mathbf{z}) = E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond \mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] \quad (3.191)$$

$$= E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] \mathbf{H}^* \quad (3.192)$$

$$= \mathbf{P} \mathbf{H}^* \quad (3.193)$$

where the second line follows from Lemma 43 given on page 3-26. Substituting the expressions in Equations (3.177) and (3.193) into Equation (3.156) yields

$$\hat{\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{x}} + \mathbf{P} \mathbf{H}^* [\mathbf{H} \mathbf{P} \mathbf{H}^* + \mathbf{R}]^{-1} (\mathbf{z} - \mathbf{H} \boldsymbol{\mu}_{\mathbf{x}}) \quad (3.183)$$

Next we'll show the error covariance operator given in Equation (3.184). We've already determined $\Sigma(\mathbf{x}, \mathbf{z})$ and $\Sigma(\mathbf{z})$ per the assumptions for this theorem. Now, substituting $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$ and $\boldsymbol{\mu}_{\mathbf{z}} = \mathbf{H}\boldsymbol{\mu}_{\mathbf{x}}$ into the definition for cross-covariance $\Sigma(\mathbf{z}, \mathbf{x})$ and then rearranging as necessary yields

$$\Sigma(\mathbf{z}, \mathbf{x}) \triangleq E[(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] \quad (3.194)$$

$$= E[(\mathbf{H}\mathbf{x} + \mathbf{v} - \mathbf{H}\boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] \quad (3.195)$$

$$= E\{[\mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) + \mathbf{v}] \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})\} \quad (3.196)$$

$$= E\{[\mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + [\mathbf{v} \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})]\} \quad (3.197)$$

¹⁸See Lemma 73, page 3-58.

Then applying the expectation operator to both terms gives

$$\Sigma(\mathbf{z}, \mathbf{x}) = E[\mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + E[\mathbf{v} \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] \quad (3.198)$$

$$= \mathbf{H}E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \diamond (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})] + \Sigma(\mathbf{v}, \mathbf{x}) \quad (3.199)$$

$$= \mathbf{H}\mathbf{P} \quad (3.200)$$

where we note that $\Sigma(\mathbf{v}, \mathbf{x}) = \mathbf{0}$ since \mathbf{v} and \mathbf{x} are independent.

Finally, substituting the covariances found in Equations (3.161), (3.193), (3.177), and (3.200) into

$$\Sigma(\mathbf{e}) = \Sigma(\mathbf{x}) - \mathbf{K} \Sigma(\mathbf{z}, \mathbf{x}) = \Sigma(\mathbf{x}) - \Sigma(\mathbf{x}, \mathbf{z}) \Sigma^{-1}(\mathbf{z}) \Sigma(\mathbf{z}, \mathbf{x}) \quad (3.125)$$

yields

$$\Sigma(\mathbf{e}) = \mathbf{P} - \mathbf{P}\mathbf{H}^*[\mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}]^{-1}\mathbf{H}\mathbf{P} \quad (3.184)$$

Q.E.D. ■

The operation of a Kalman filter is a natural two-step recursive process, consisting of a state update with the latest measurement and a state prediction based on the dynamics model. In the final theorem of this section, we present the LIMVUE for a stochastic state process using a stochastic measurement process. We will accomplish this by generalizing the LIMVUE, given in Theorem 75, for a stochastic measurement process which is a generalization of Definition 72.

Definition 76 (Generalized Linear Stochastic Measurement Model) *Let \mathbf{z} , \mathbf{x} , and \mathbf{v} be discrete-time stochastic processes which map the product space $\mathbb{T} \times \Omega$ into their respective realization spaces \mathbb{Z} , \mathbb{X} , and \mathbb{V} , where $\mathbb{T} \subset \mathbb{R}^+$ and Ω is the sample space, a nonempty set associated with a complete probability space (Ω, \mathcal{F}, P) . At each time $t_i \in \mathbb{T}$, $\mathbf{z}(t_i) \equiv \mathbf{z}(t_i, \cdot) \in \tilde{\mathbb{Z}}$, $\mathbf{x}(t_i) \equiv \mathbf{x}(t_i, \cdot) \in \tilde{\mathbb{X}}$, and $\mathbf{v}(t_i) \equiv \mathbf{v}(t_i, \cdot) \in \tilde{\mathbb{V}}$*

are random vectors. The measurement process model is defined by

$$\mathbf{z}(t_i) = \mathbf{H}(t_i) \mathbf{x}(t_i) + \mathbf{v}(t_i) \quad (3.201)$$

where:

$$\begin{aligned} \mathbf{z}(t_i) &\in \tilde{\mathbb{Z}} = \mathbb{L}^2(\Omega, P; \mathbb{Z}) & \dots & \text{measurement vector} \\ \mathbf{H}(t_i) &\in \mathcal{BLT}(\tilde{\mathbb{X}}, \tilde{\mathbb{Z}}) & \dots & \text{measurement distributor transformation} \\ \mathbf{x}(t_i) &\in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X}) & \dots & \text{state vector} \\ \mathbf{v}(t_i) &\in \tilde{\mathbb{V}} = \mathbb{L}^2(\Omega, P; \mathbb{V}) & \dots & \text{measurement-corruption noise vector} \end{aligned}$$

Additionally, $\mathbf{H}(t_i)$ is known and the white measurement-corruption noise and state covariances are defined as

$$\Sigma[\mathbf{v}(t_i), \mathbf{v}(t_j)] = E[\mathbf{v}(t_i) \diamond \mathbf{v}(t_j)] = \mathbf{R}(t_i) \delta_{ij} \quad (3.202)$$

$$\Sigma[\mathbf{x}(t_i), \mathbf{x}(t_i)] = E\{[\mathbf{x}(t_i) - \boldsymbol{\mu}_{\mathbf{x}}(t_i)] \diamond [\mathbf{x}(t_i) - \boldsymbol{\mu}_{\mathbf{x}}(t_i)]\} = \mathbf{P}(t_i) \quad (3.203)$$

Since $\mathbf{v}(t_i)$ and $\mathbf{x}(t_j)$ are independent for all times t_i and t_j , and $E[\mathbf{v}(t_i)] = \mathbf{0}$, the cross-correlations and cross-covariances are zero. Per Definition 45, the state covariance operator, $\mathbf{P}(t_i)$, is symmetric, positive, and nuclear for all time t_i , whereas the measurement-corruption noise covariance operator, $\mathbf{R}(t_i)$, is symmetric, strictly positive, and nuclear for all time t_i .

The boxology for this generalized stochastic measurement model is given in Figure 3.10.

While the measurement history, defined in Equation (2.18) on page 2-14, “stored” the elements in a growing vector, the following definition uses a set because the measurement vector may be infinite-dimensional, hence the measurement history may be infinite-dimensional.

Definition 77 (Measurement History) *Let $\mathbf{z}(\cdot, \cdot)$ be a discrete-time stochastic measurement process in accordance with the generalized linear stochastic measure-*

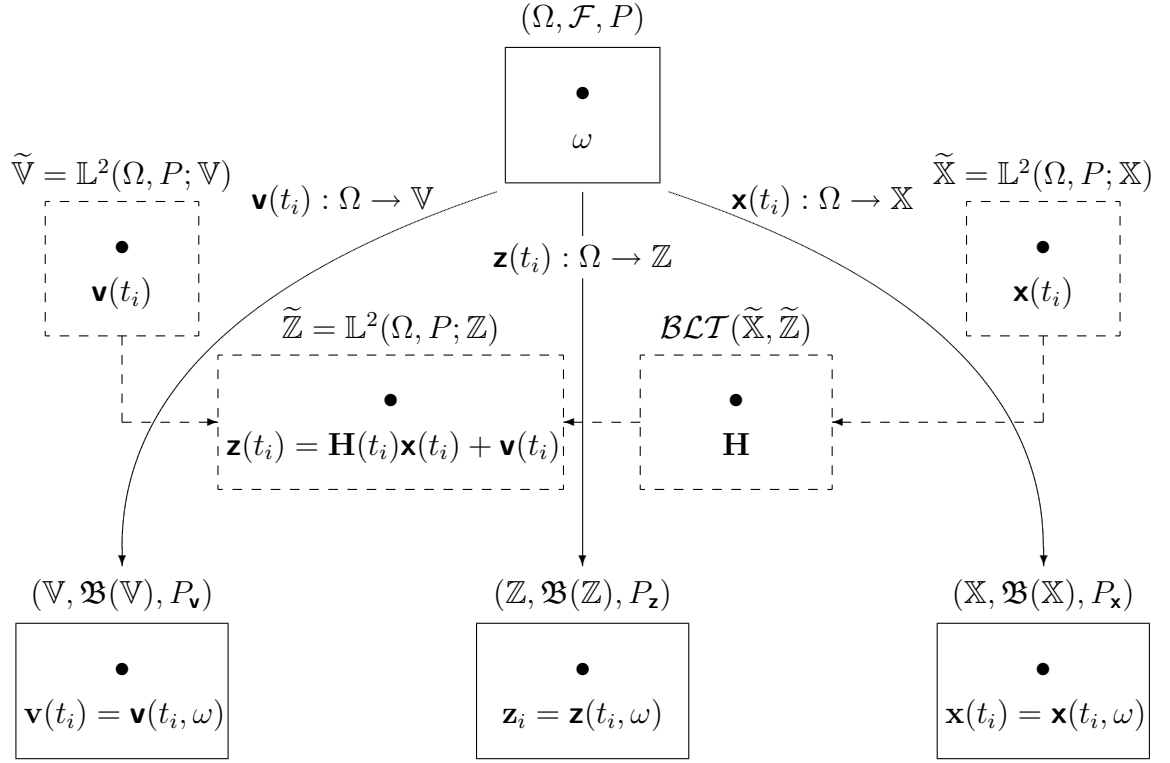


Figure 3.10 Boxology of a Generalized Linear Stochastic Measurement Model.

ment model of Definition 76. The stochastic measurement history and measurement history sample, through time t_i are defined as

$$\mathbf{Z}(t_i) \triangleq \{\mathbf{z}(t_1), \mathbf{z}(t_2), \dots, \mathbf{z}(t_i)\} \quad (3.204)$$

$$\mathbf{Z}_i \triangleq \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_i\} \quad (3.205)$$

respectively, where \mathbf{z}_i is a convenient notation for $\mathbf{z}(t_i)$, a specific realization of the random vector $\mathbf{z}(t_i)$. Additionally, since the sets “grow” with each new measurement, they are related as

$$\mathbf{Z}(t_1) \subset \mathbf{Z}(t_2) \subset \dots \subset \mathbf{Z}(t_i) \subset \tilde{\mathbb{Z}} \quad (3.206)$$

$$\mathbf{Z}_1 \subset \mathbf{Z}_2 \subset \dots \subset \mathbf{Z}_i \subset \mathbb{Z} \quad (3.207)$$

Now we can state (in a theorem) the LIMVUE for a stochastic measurement process — an estimator that employs a sequence of measurements to improve the existing estimate. By this, we mean an estimator $\hat{\mathbf{x}}(t_{i+1})$ that *updates* $\hat{\mathbf{x}}(t_i)$ using the new measurement $\mathbf{z}(t_{i+1})$ — a recursive estimator for a stochastic process.

Theorem 78 (LIMVUE for Stochastic Processes) *Let the state, \mathbf{x} , measurement, \mathbf{z} , and measurement-corruption noise, \mathbf{v} , be as described in Definition 76 and the measurement history in Definition 77. Let $\mathbf{z}(t_j)$ for $j = 1, 2, \dots, i$ be the random measurement vectors generating a subspace $\tilde{\mathbb{Z}}_i$ of $\tilde{\mathbb{Z}}$. Then $\hat{\mathbf{x}}(t_i) \triangleq \mathcal{E}[\mathbf{x}(t_i)|\mathbf{Z}(t_i)]$ is the conditional state estimator, an orthogonal projection on closed subspace $\tilde{\mathbb{Z}}_i$ of $\tilde{\mathbb{Z}}$. Note that expectations seen in Theorems 71 and 75 are all now replaced with conditional expectations, conditioned on the previous measurement history. Finally, we denote the projection of $\mathbf{z}(t_{i+1})$ onto subspace $\tilde{\mathbb{Z}}_i$ by $\hat{\mathbf{z}}(t_{i+1}^-)$, where $\mathbf{z}(t_{i+1})$ generates the subspace $\tilde{\mathbb{Z}}_{i+1}$ and the superscript “ $-$ ” indicates that the estimate is based on the “old” information up through time t_i .*

Therefore, the LIMVUE for a stochastic measurement process is the conditional state estimator

$$\hat{\mathbf{x}}(t_i) = \hat{\mathbf{x}}(t_i^-) + \mathbf{K}(t_i)[\mathbf{z}(t_i) - \hat{\mathbf{z}}(t_i^-)] \quad (3.208)$$

where $\hat{\mathbf{x}}(t_i^-)$ is the conditional state estimator based on the “old” information up through time t_i , that is, $\hat{\mathbf{x}}(t_i^-) \triangleq \mathcal{E}[\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})]$, and the Kalman gain transformation takes the form

$$\mathbf{K}(t_i) = \mathbf{P}(t_i)\mathbf{H}^*(t_i)\mathbf{A}^{-1}(t_i) \quad (3.209)$$

and the filter-computed residual covariance operator is

$$\mathbf{A}(t_i) \equiv \mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}^*(t_i) + \mathbf{R}(t_i) \quad (3.210)$$

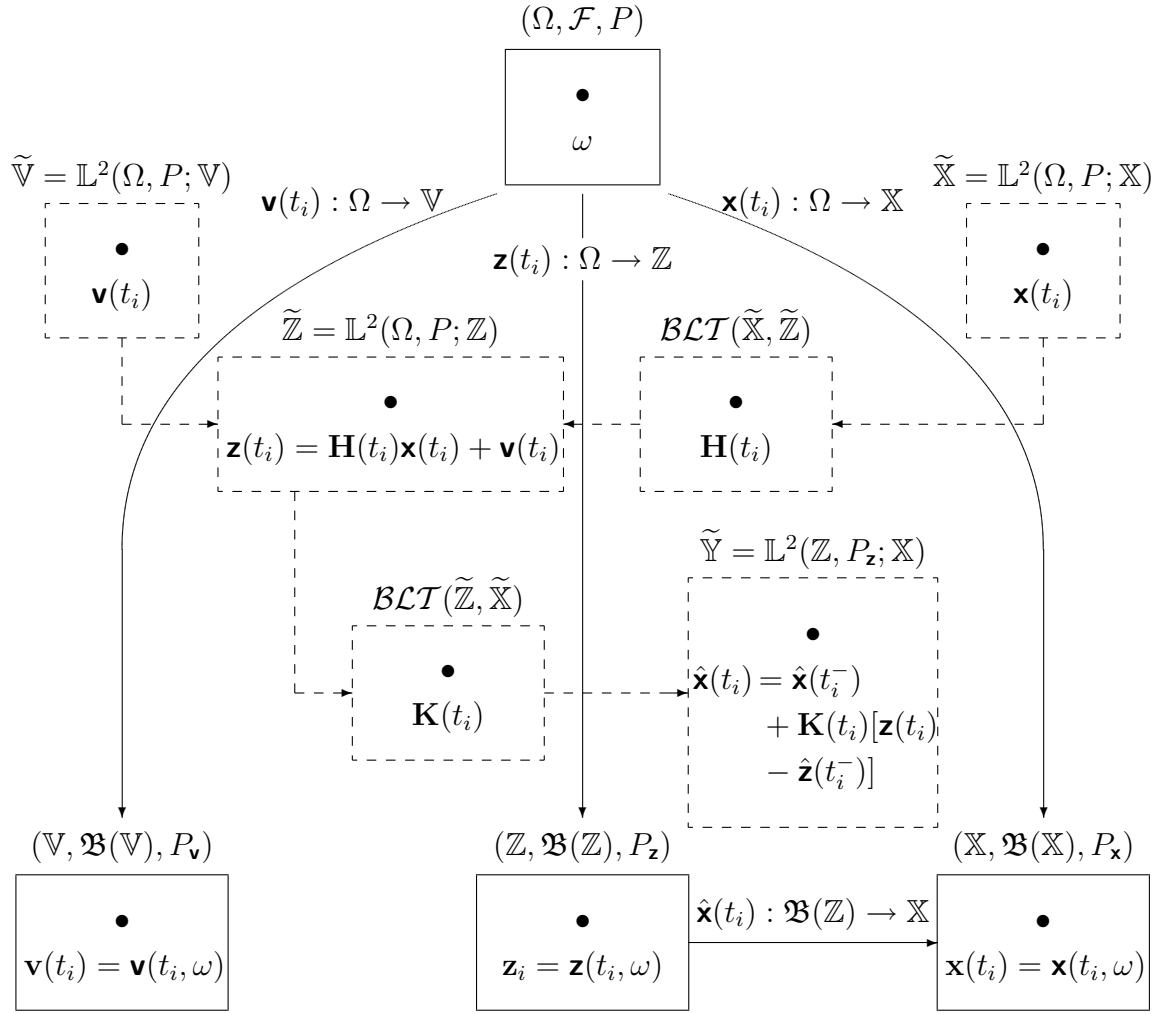


Figure 3.11 Boxology of the Stochastic LIMVUE.

Additionally, the corresponding conditional error covariance operator, defined by,

$$\begin{aligned}
 \mathbf{P}(t_i) &\triangleq \Sigma\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i)] | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \\
 &= E\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i)] \diamond [\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i)] | \mathbf{Z}(t_i) = \mathbf{Z}_i\}
 \end{aligned} \tag{3.211}$$

is given by

$$\mathbf{P}(t_i) = \mathbf{P}(t_i^-) - \mathbf{K}(t_i)\mathbf{H}(t_i)\mathbf{P}(t_i^-) \tag{3.212}$$

The boxology for this estimator is given in Figure 3.11.

Remark Thus far, we have only considered the measurement side in preparation of constructing the ISKF, without due regard to the dynamics of the system (or process); this will be addressed in the following section when we introduce the dynamics model for the system. After the dynamics model is added, Equation (3.208) will be amended to reflect that we are now updating the “propagated” state estimator versus the previous state estimator, $\hat{\mathbf{x}}(t_1)$, based on the measurement at time t_1 . The “propagated” state estimator is due to the dynamics model “propagating” the state estimator from time t_1 to time t_2 ; we denote this new estimator by $\hat{\mathbf{x}}(t_2^-)$, where the superscript minus sign indicates the time instant just prior to measurement incorporation.

Proof of Theorem 78 If we make the following substitutions:

$$\mathbf{z} = \mathbf{z}(t_i) \quad (3.213)$$

$$\boldsymbol{\mu}_{\mathbf{z}} = \hat{\mathbf{z}}(t_i^-) \quad (3.214)$$

$$\mathbf{x} = \mathbf{x}(t_i) \quad (3.215)$$

$$\boldsymbol{\mu}_{\mathbf{x}} = \hat{\mathbf{x}}(t_i^-) \quad (3.216)$$

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}(t_i) \quad (3.217)$$

then we have restated Theorem 71, and thus with the aid of Lemma 74, Equation (3.208) follows and therefore Theorem 78 holds. Note that we have used means conditioned on the previous measurement history, $\hat{\mathbf{z}}(t_i^-)$ and $\hat{\mathbf{x}}(t_i^-)$, in the place of unconditional means, $\boldsymbol{\mu}_{\mathbf{z}}$ and $\boldsymbol{\mu}_{\mathbf{x}}$, in our substitution scheme, respectively. ■

Another way of verifying Theorem 78 begins by noting that the conditional state estimator $\hat{\mathbf{x}}(t_1)$ is the best estimate of $\mathbf{x}(t_1)$ given subspace $\tilde{\mathbb{Z}}_1$, where $\tilde{\mathbb{Z}}_1$ was generated by measurement $\mathbf{z}(t_1)$. The next measurement $\mathbf{z}(t_2)$, along with $\mathbf{z}(t_1)$, generates subspace $\tilde{\mathbb{Z}}_2$, while $\hat{\mathbf{z}}(t_2^-)$ denotes the projection of $\mathbf{z}(t_2)$ onto subspace $\tilde{\mathbb{Z}}_1$. Hence, $\hat{\mathbf{z}}(t_2^-)$ is the best estimate of $\mathbf{z}(t_2)$ given subspace $\tilde{\mathbb{Z}}_1$. Let $\mathbf{r}(t_2) = \mathbf{z}(t_2) - \hat{\mathbf{z}}(t_2^-)$ be the residual (or difference between the true measurement and the best prediction

of the measurement) at time t_2 . It then follows that the projection of $\mathbf{x}(t_2)$ onto the subspace of $\tilde{\mathbb{Z}}_2$, denoted $\hat{\mathbf{x}}(t_2)$, is given by

$$\hat{\mathbf{x}}(t_2) = \hat{\mathbf{x}}(t_1) + \Sigma[\mathbf{x}(t_2), \mathbf{r}(t_2)] \Sigma^{-1}[\mathbf{r}(t_2)] \mathbf{r}(t_2) \quad (3.218)$$

Now observe that the estimate of $\mathbf{x}(t_2)$, which lives in subspace $\tilde{\mathbb{Z}}_1 + \tilde{\mathbb{Z}}_2 \subset \tilde{\mathbb{Z}}$, can be decomposed as

$$\tilde{\mathbb{Z}}_1 + \tilde{\mathbb{Z}}_2 = \tilde{\mathbb{Z}}_1 \oplus \tilde{\mathbb{Y}}_2 \quad (3.219)$$

where $\tilde{\mathbb{Y}}_2 = \{\mathbf{z} : \mathbf{z} \in \tilde{\mathbb{Z}}_2 \text{ and } \mathbf{z} \notin \tilde{\mathbb{Z}}_1\}$, or in words, $\tilde{\mathbb{Y}}_2$ is that part of $\tilde{\mathbb{Z}}_2$, which is not in $\tilde{\mathbb{Z}}_1 \cap \tilde{\mathbb{Z}}_2$, and \oplus is the direct sum operation. Thus $\tilde{\mathbb{Y}}_2$ and $\tilde{\mathbb{Z}}_1$ are orthogonal subspaces, i.e., $\tilde{\mathbb{Y}}_2 \perp \tilde{\mathbb{Z}}_1$. Hence, Equation (3.218) can be written as

$$\hat{\mathbf{x}}(t_2) = \hat{\mathbf{x}}(t_1) + \mathbf{y}(t_2) \quad (3.220)$$

where

$$\mathbf{y}(t_2) = \Sigma[\mathbf{x}(t_2), \mathbf{r}(t_2)] \Sigma^{-1}[\mathbf{r}(t_2)] \mathbf{r}(t_2) \in \tilde{\mathbb{Y}}_2 \quad (3.221)$$

represents the new information about \mathbf{x} brought by the second measurement and

$$\hat{\mathbf{x}}(t_1) \in \tilde{\mathbb{Z}}_1 \quad (3.222)$$

represents the best estimate from the first (or previous) measurement. Therefore, we see that the projection onto a sum of subspaces is equal to the sum of individual projections when the subspaces are orthogonal. Q.E.D. again.

3.4 Dynamics Model

One of the first tasks in model-based estimation is to create a mathematical model of the system of interest. Many physically motivated problems are well mod-

eled using a linear continuous-time dynamics model of the stochastic system; our first definition is an abstraction of that model.

Definition 79 (Continuous-Time Dynamics Model) *A continuous-time and space model for the linear dynamics of a stochastic state process \mathbf{x} can be viewed as a set of random vectors $\{\mathbf{x}(t) : t \in \mathbb{T}\}$, where $\mathbb{T} = [t_0, t_f] \subset \mathbb{R}^+$, by the stochastic differential equation*

$$\begin{aligned} d\mathbf{x}(t) &= [\mathbf{F}(t) \mathbf{x}(t) + \mathbf{B}(t) \mathbf{u}(t)]dt + \mathbf{G}(t) d\mathbf{b}(t) \\ \mathbf{x}(t_0) &= \mathbf{x}_0 \end{aligned} \tag{3.223}$$

where the vectors, operators, and transformations are defined at time t as

$$\begin{aligned} \mathbf{x}(t) &\in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X}) && \dots && \text{state vector} \\ \mathbf{F}(t) &\in \mathcal{LO}(\tilde{\mathbb{X}}) && \dots && \text{state distributor operator} \\ \mathbf{B}(t) &\in \mathcal{LT}(\mathbb{U}, \tilde{\mathbb{X}}) && \dots && \text{input distributor transformation} \\ \mathbf{u}(t) &\in \mathbb{U} && \dots && \text{known input vector} \\ \mathbf{G}(t) &\in \mathcal{BLT}(\tilde{\mathbb{B}}, \tilde{\mathbb{X}}) && \dots && \text{noise distributor transformation} \\ \mathbf{b}(t) &\in \tilde{\mathbb{B}} = \mathbb{L}^2(\Omega, P; \mathbb{B}) && \dots && \text{Brownian motion noise vector} \end{aligned}$$

Additionally, \mathbf{b} is a Brownian motion process (and thus an independent increment process¹⁹) with constant diffusion operator \mathbf{Q} as discussed in Definition 61. Furthermore, the dynamics model must include the pertinent boundary conditions for the specific problem at hand. For example, if $\mathbf{F} = \nabla$, a gradient, then we could associate a Dirichlet or Neumann boundary condition with Equation (3.223).

While many of the problems of interest are best modeled with a continuous-time description discussed in Definition 79, we will most likely need a discrete-time model so that the eventual filtering algorithm can be implemented on a digital computer as software. On the other hand, if the problem is posed in a discrete-time format, then the following definition still applies, but without the interpretation

¹⁹The most general form of additive noise need only be an “independent increment” process, hence we may also develop a continuous-time dynamics model which features a generalized Poisson noise process [141, 129, 66].

from the continuous-time model, as is already stated. One drawback to a naturally discrete-time dynamics model is that there is no guarantee that the state transition operator, as defined below, will be invertible; this is precisely the same situation as discussed in Section 2.3.1, regarding the state transition matrix for the finite-dimensional dynamics model.

The following model is termed an *equivalent* discrete-time model²⁰ since the state at any time t_i is precisely the same as the state using the continuous-time model at time $t = t_i$. Note that the set of time instants in the continuous-time model is a continuum, i.e., $\mathbb{T} = [t_0, t_f]$, whereas for the discrete-time model, it is a discrete set: $\mathbb{T} = \{t_0, t_1, \dots, t_f\}$. The process for creating the equivalent discrete-time model is a substantially different process than merely sampling the continuous-time process, as will be seen in the development following the definition.

Definition 80 (Discrete-Time Dynamics Model) *A discrete-time and space model for the linear dynamics of a stochastic state process \mathbf{x} can be viewed as a sequence of random vectors $\{\mathbf{x}(t) : t \in \mathbb{T}\}$, where $\mathbb{T} = \{t_0, t_1, \dots, t_f\} \subset \mathbb{R}^+$, by the stochastic difference equation*

$$\begin{aligned} \mathbf{x}(t_{i+1}) &= \Phi(t_{i+1}, t_i) \mathbf{x}(t_i) + \mathbf{B}_d(t_i) \mathbf{u}(t_i) + \mathbf{G}_d(t_i) \mathbf{w}_d(t_i) \\ \mathbf{x}(t_0) &= \mathbf{x}_0 \end{aligned} \tag{3.224}$$

where the vectors, operators, and transformations are defined at time t_i as

$$\begin{aligned} \mathbf{x}(t_i) &\in \widetilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X}) & \dots & \text{state vector} \\ \Phi(t_{i+1}, t_i) &\in \mathcal{BLO}(\widetilde{\mathbb{X}}) & \dots & \text{state transition operator from time } t_i \text{ to } t_{i+1} \\ \mathbf{B}_d(t_i) &\in \mathcal{LT}(\mathbb{U}, \widetilde{\mathbb{X}}) & \dots & \text{discrete-time input distributor transformation} \\ \mathbf{u}(t_i) &\in \mathbb{U} & \dots & \text{known control input vector} \\ \mathbf{G}_d(t_i) &\in \mathcal{BLT}(\widetilde{\mathbb{W}}, \widetilde{\mathbb{X}}) & \dots & \text{discrete-time noise distributor transformation} \\ \mathbf{w}_d(t_i) &\in \widetilde{\mathbb{W}} = \mathbb{L}^2(\Omega, P; \mathbb{W}) & \dots & \text{zero-mean white Gaussian noise vector} \end{aligned}$$

²⁰See Maybeck [129] for the analogous finite-dimensional case.

where the zero-mean white Gaussian dynamics noise process, $\mathbf{w}_d(\cdot, \cdot)$, has covariance kernel

$$\Sigma[\mathbf{w}_d(t_i), \mathbf{w}_d(t_j)] \triangleq E[\mathbf{w}_d(t_i) \diamond \mathbf{w}_d(t_j)] = \mathbf{Q}_d(t_i) \delta_{ij} \quad (3.225)$$

where the bounded and linear covariance operator $\mathbf{Q}_d(t_i) \in \mathcal{BLO}(\widetilde{\mathbb{W}})$ is symmetric, positive, and nuclear for all time t_i . Furthermore, the initial state condition $\mathbf{x}(t_0)$ is not known precisely; it will be modeled as a Gaussian random vector, independent of $\mathbf{w}_d(\cdot, \cdot)$ with mean and covariance specified as follows

$$E[\mathbf{x}(t_0)] = \hat{\mathbf{x}}_0 \quad (3.226)$$

$$\Sigma[\mathbf{x}(t_0)] \triangleq E\{[\mathbf{x}(t_0) - \hat{\mathbf{x}}_0] \diamond [\mathbf{x}(t_0) - \hat{\mathbf{x}}_0]\} = \mathbf{P}_0 \quad (3.227)$$

where the initial error covariance operator, $\mathbf{P}_0 \in \mathcal{BLO}(\widetilde{\mathbb{X}})$, is symmetric, positive, and nuclear. For time $t_{i+1} = t_1, \dots, t_f$, the conditional error covariance operator is defined by

$$\begin{aligned} \mathbf{P}(t_{i+1}^-) &\triangleq \Sigma\{[\mathbf{x}(t_{i+1}) - \hat{\mathbf{x}}(t_{i+1})] | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \\ &= E\{[\mathbf{x}(t_{i+1}) - \hat{\mathbf{x}}(t_{i+1})] \diamond [\mathbf{x}(t_{i+1}) - \hat{\mathbf{x}}(t_{i+1})] | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \end{aligned} \quad (3.228)$$

The boxology for the discrete-time dynamics model appears in Figure 3.12. Note that, for illustrative purposes, we have chosen to use two boxes to represent the Hilbert space of random vectors, $\widetilde{\mathbb{X}}$, one for time t_i and one for time t_{i+1} . On the other hand, we have just one box for the realization space, $(\mathbb{X}, \mathfrak{B}(\mathbb{X}), P_{\mathbf{x}})$.

If the discrete-time dynamics model is based on the continuous-time model detailed in Definition 79, then $\Phi(t_{i+1}, t_i)$, $\mathbf{B}_d(t_i)$, $\mathbf{u}(t_i)$, $\mathbf{G}_d(t_i)$, $\mathbf{Q}_d(t_i)$, and $\mathbf{w}_d(t_i)$ are based on the following development entailing a series of definitions and theorems, which results from solving the stochastic differential equation. The solution (or mild

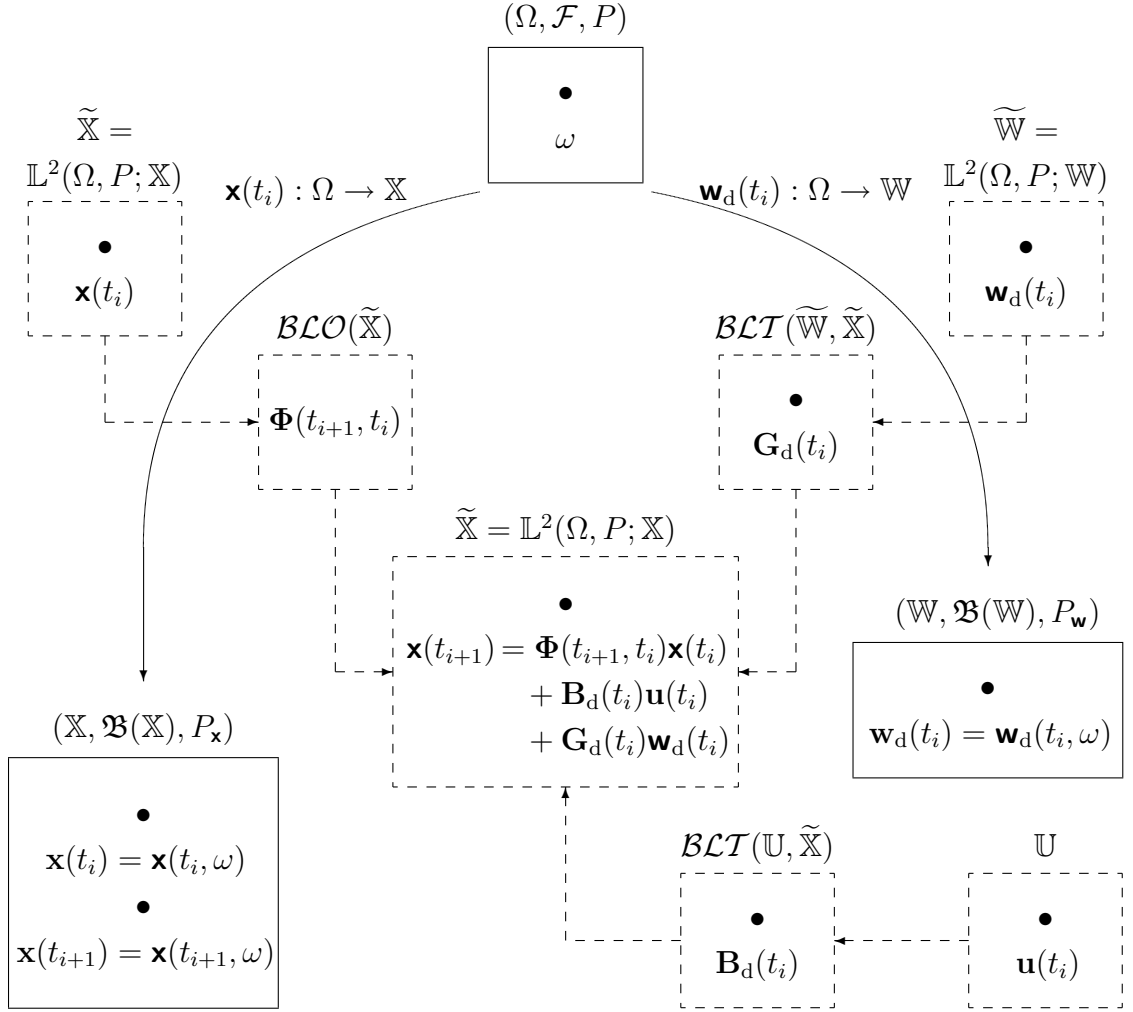


Figure 3.12 Boxology of a Discrete-Time Dynamics Model.

form) of Equation (3.223) is given by the so called *evolution* system [38, 66]

$$\mathbf{x}(t) = \Phi(t, t_0) \mathbf{x}_0 + \int_{t_0}^t \Phi(t, s) \mathbf{B}(s) \mathbf{u}(s) ds + \int_{t_0}^t \Phi(t, s) \mathbf{G}(s) d\mathbf{b}(s) \quad (3.229)$$

where $\Phi(t, s)$ is the state transition or mild evolution operator associated with the state distributor operator, $\mathbf{F}(t)$ [38, 160]. The theory for evolution operators is rather technical and is not needed to develop the theory for the class of problems

in this research²¹. Therefore, the full theory for (one- and two-parameter) evolution operators will be neither reviewed nor pursued in this research.

While a time-varying state distributor operator, $\mathbf{F}(t)$, generates a semigroup of *two*-parameter state transition operators, $\Phi(t, s)$, a time-invariant state distributor operator, \mathbf{F} , generates a semigroup of *one*-parameter state transition operators, $\Phi(t - s)$ [38, 160, 39, 48, 115]. The single parameter is denoted by the “time” difference $t - s$ for $0 \leq s < t < \infty$, just as it was for the finite-dimensional case we reviewed in Section 2.3.1, Equation (2.8), on page 2-11.

Briefly, the plan for the rest of this section falls into two main parts. First, we will discuss two types of one-parameter semigroups of BLOs. Then, we will match up the terms in Equations (3.223) and (3.229) as we determine \mathbf{B}_d , \mathbf{G}_d , and \mathbf{w}_d .

The pertinent theory for generating these one-parameter semigroups is included in the following series of definitions and theorems. We begin with the definition for a semigroup of BLOs.

Definition 81 (Semigroup of BLOs) *Let \mathbb{X} be a Banach space. A one-parameter family of BLOs, denoted by $\{\Phi(t) : t \geq 0\}$ for $0 \leq t < \infty$, is a semigroup of BLOs on \mathbb{X} if [160]:*

1. $\Phi(0) = \mathbf{I}$, where \mathbf{I} is the identity operator on \mathbb{X} , and
2. $\Phi(t + s) = \Phi(t)\Phi(s)$ for every $t, s \geq 0$

There are several types of one-parameter semigroups; we shall discuss just two: the uniformly and the strongly continuous semigroup of BLOs. These categories of operators are due to the nature of the generating time-invariant operator \mathbf{F} discussed above. The uniformly continuous semigroup of BLOs is included to show where the

²¹Limiting our explication to one-parameter semigroups is not as restrictive as it may seem according to Engel and Nagel [48]. They (along with their collaborators) have studied population, nuclear transport, delay differential, and Volterra equations, and both ordinary and partial differential operators in the form of an abstract Cauchy problem using the theory of one-parameter semigroups.

finite-dimensional theory and the infinite-dimensional theory appear to agree in the form of their equations, while the strongly continuous semigroup of BLOs is needed for our extended example discussed in the next chapter.

Definition 82 (Uniformly Continuous Semigroup of BLOs) *A semigroup of BLOs, $\{\Phi(t) : t \geq 0\}$, is said to be uniformly continuous if*

$$\lim_{t \downarrow 0} \|\Phi(t) - \mathbf{I}\| = 0 \quad (3.230)$$

In general, \mathbf{F} does not need to be bounded, as seen in Definition 79. However, when \mathbf{F} is bounded, it is the infinitesimal generator for a uniformly continuous semigroup of operators.

Theorem 83 (Properties of a Uniformly Continuous Semigroup of BLOs)

Let $\{\Phi(t) : t \geq 0\}$ be a uniformly continuous semigroup of BLOs. The infinitesimal generator \mathbf{F} for a uniformly continuous semigroup is a BLO. Then [160]:

1. *There exists a constant $\omega \geq 0$ such that $\|\Phi(t)\| \leq \exp(\omega t)$.*
2. *There exists a unique BLO \mathbf{F} such that $\Phi(t) = \exp(t\mathbf{F})$.*
3. *The operator \mathbf{F} in part (2) is the infinitesimal generator of $\{\Phi(t) : t \geq 0\}$.*
4. *$t \mapsto \Phi(t)$ is differentiable in norm and*

$$\frac{d\Phi(t)}{dt} = \mathbf{F}\Phi(t) = \Phi(t)\mathbf{F} \quad (3.231)$$

Proof of Theorem 83 See Pazy [160].

In our research, our \mathbf{F} is unbounded. Thus, we employ the strongly continuous semigroup of BLOs, which is defined next.

Definition 84 (Strongly Continuous Semigroup of BLOs) *Let \mathbb{X} be a Banach space. A semigroup, $\{\Phi(t) : t \geq 0\}$, of BLOs on \mathbb{X} is a strongly continuous semigroup of BLOs, for $0 \leq t < \infty$, if*

$$\lim_{t \downarrow 0} \Phi(t) \mathbf{x} = \mathbf{x} \quad (3.232)$$

for every \mathbf{x} in \mathbb{X} . A strongly continuous semigroup of BLOs on \mathbb{X} is called a semigroup of class C_0 or simply a C_0 semigroup.

Some useful properties of this class of semigroup operators are reported in the following theorem²².

Theorem 85 (Properties of a Strongly Continuous Semigroup of BLOs)

Let $\Phi(t)$ be a C_0 semigroup and let \mathbf{F} be the infinitesimal generator. Then [160]:

1. *For $\mathbf{x} \in \mathbb{X}$*

$$\lim_{h \rightarrow 0} \frac{1}{h} \int_t^{t+h} \Phi(s) \mathbf{x} ds = \Phi(t) \mathbf{x} \quad (3.233)$$

2. *For $\mathbf{x} \in \mathbb{X}$*

$$\int_0^t \Phi(s) \mathbf{x} ds \in \mathcal{D}(\mathbf{F}) \quad (3.234)$$

where $\mathcal{D}(\mathbf{F})$ denotes the domain and

$$\mathbf{F} \left(\int_0^t \Phi(s) \mathbf{x} ds \right) = \Phi(t) \mathbf{x} - \mathbf{x} \quad (3.235)$$

3. *For $\mathbf{x} \in \mathcal{D}(\mathbf{F})$,*

$$\Phi(t) \mathbf{x} - \Phi(s) \mathbf{x} = \int_s^t \Phi(\tau) \mathbf{F} \mathbf{x} d\tau = \int_s^t \mathbf{F} \Phi(\tau) \mathbf{x} d\tau \quad (3.236)$$

²²Curtain [38, 39] and Lax [115] also present a similar collection of semigroup properties in the infinite-dimensional linear systems framework.

4. For $\mathbf{x} \in \mathcal{D}(\mathbf{F})$, we have $\Phi(t) \mathbf{x} \in \mathcal{D}(\mathbf{F})$ and

$$\frac{d}{dt} \Phi(t) \mathbf{x} = \mathbf{F} \Phi(t) \mathbf{x} = \Phi(t) \mathbf{F} \mathbf{x} \quad (3.237)$$

Proof of Theorem 85 See Pazy [160].

Now that we've laid the groundwork for the one-parameter semigroup of BLOs, we turn to finding the equivalent discrete-time transformations and additive noise shown in Equation (3.224). While we have generally limited our discussion of the semigroups to those with a single parameter, the theory that follows does not depend on this simplification; hence, the remainder of the development will employ the (less restrictive) two-parameter notation for the state transition operator, $\Phi(t, s)$.

Theorem 86 (Equivalent Discrete-Time Input Distributor Transformation) *Given the continuous-time dynamics model in Definition 79 and the desired form of the discrete-time dynamics model in Definition 80, the equivalent discrete-time input distributor transformation is given by*

$$\mathbf{B}_d(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, \tau) \mathbf{B}(\tau) d\tau \quad (3.238)$$

provided $\mathbf{u}(t)$ is a piece-wise constant function, constant over each sample period.

Remarks (1) Practically speaking, the restriction on control input $\mathbf{u}(t)$ merely reflects that case for adjusting the control input at the end of each propagation cycle, i.e., after a sample period has ended. If this control input is generated by a digital computer, the interface to the continuous-time system is assumed to be through a zero-order hold, thus keeping the control value constant over the ensuing sample period. (2) For many problems, such as the heat equation example discussed in the following chapter, the control input may also be a function of the spatial dimension. There is nothing in this definition that precludes us from allowing the control input to vary over the spatial dimension.

Proof of Theorem 86 If we were to compute the state using Equation (3.229) at the discrete time instants $\mathbb{T} = \{t_0, t_1, \dots, t_f\}$, then we could rewrite it with subscripted time arguments as follows

$$\begin{aligned} \mathbf{x}(t_{i+1}) = & \Phi(t_{i+1}, t_i) \mathbf{x}(t_i) + \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s) \mathbf{B}(s) \mathbf{u}(s) ds \\ & + \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s) \mathbf{G}(s) d\mathbf{b}(s) \end{aligned} \quad (3.239)$$

where $t_0 \leq t_i < t_{i+1} \leq t_f$. Matching up the terms in Equations (3.224) and (3.239) yields

$$\mathbf{B}_d(t_i) \mathbf{u}(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s) \mathbf{B}(s) \mathbf{u}(s) ds \quad (3.240)$$

If we assume that the control input is constant for this particular time interval, i.e., $\mathbf{u}(s) = \mathbf{u}(t_i)$ for $t_i \leq s \leq t_{i+1}$, then we may pull $\mathbf{u}(s)$ outside of the integral in Equation (3.240) to get

$$\mathbf{B}_d(t_i) \mathbf{u}(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s) \mathbf{B}(s) ds \mathbf{u}(t_i) \quad (3.241)$$

Since Equation (3.241) holds for all piece-wise constant $\mathbf{u}(t)$, we have found the $\mathbf{B}_d(t_i)$ operator in Equation (3.238). ■

Definition 87 (Equivalent Discrete-Time Noise Distributor Transformation) *Analogous to the finite-dimensional case developed in [129], we define $\mathbf{G}_d(t_i)$ to be an identity operator on $\widetilde{\mathbb{W}}$.*

Theorem 88 (Equivalent Discrete-Time Noise Characterization) *For the continuous-time dynamics model in Definition 79 and the desired form of the discrete-time dynamics model in Definition 80, the equivalent discrete-time noise vector, defined by*

$$\mathbf{w}_d(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t, s) \mathbf{G}(s) d\mathbf{b}(s) \quad (3.242)$$

is a zero-mean white Gaussian with covariance kernel operator, $\Sigma[\mathbf{w}_d(t_i), \mathbf{w}_d(t_k)] = \mathbf{0}$, for $t_i \neq t_k$, and covariance operator

$$\Sigma[\mathbf{w}_d(t_i), \mathbf{w}_d(t_i)] = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s) \mathbf{G}(s) \mathbf{Q} \mathbf{G}^*(s) \Phi^*(t_{i+1}, s) ds = \mathbf{Q}_d(t_i) \quad (3.243)$$

whenever $t_i = t_k$.

Proof of Theorem 88 Let the third term on the right-hand side of Equation (3.239) be identified as equivalent discrete-time noise vector. To include the possibility of considering \mathbf{w}_d for more than a single interval, such as t_i to t_{i+1} , we write it as a function of two (possibly nonconsecutive) time instants $t_i < t_j$

$$\mathbf{w}_d(t_i, t_j) \triangleq \int_{t_i}^{t_j} \Phi(t_j, s) \mathbf{G}(s) d\mathbf{b}(s) \quad (3.244)$$

The mean of $\mathbf{w}_d(t_i, t_j)$ is zero, i.e., [38]

$$E[\mathbf{w}_d(t_i, t_j)] = E \left[\int_{t_i}^{t_j} \Phi(t_j, s) \mathbf{G}(s) d\mathbf{b}(s) \right] = \mathbf{0} \quad (3.245)$$

Next, we write the covariance for the case of overlapping intervals [38]

$$\Sigma[\mathbf{w}_d(t_i, t_j), \mathbf{w}_d(t_k, t_l)] \triangleq E[\mathbf{w}_d(t_i, t_j) \diamond \mathbf{w}_d(t_k, t_l)] \quad (3.246)$$

Then,

$$\begin{aligned} E[\mathbf{w}_d(t_i, t_j) \diamond \mathbf{w}_d(t_k, t_l)] \\ = \int_{\max(t_i, t_k)}^{\min(t_j, t_l)} \Phi(\min(t_j, t_l), s) \mathbf{G}(s) \mathbf{Q} \mathbf{G}^*(s) \Phi^*(\min(t_j, t_l), s) ds \end{aligned} \quad (3.247)$$

where $t_j > t_k$. The covariance for non-overlapping intervals, i.e., whenever either $t_j < t_k$ or $t_l < t_i$ is true, is [38]

$$\Sigma[\mathbf{w}_d(t_i, t_j), \mathbf{w}_d(t_k, t_l)] = \mathbf{0} \quad (3.248)$$

Note that we only assumed that $t_i < t_j$ and $t_k < t_l$, and that in general, $t_j - t_i$ need not equal $t_l - t_k$. When t_i and t_j are consecutive times such that $t_j = t_{i+1}$ and similarly for t_k and t_l , then we need only one time argument for \mathbf{w}_d ; thus, we obtain

$$\Sigma[\mathbf{w}_d(t_i), \mathbf{w}_d(t_k)] = \begin{cases} \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s) \mathbf{G}(s) \mathbf{Q} \mathbf{G}^*(s) \Phi^*(t_{i+1}, s) ds, & t_i = t_k \\ \mathbf{0}, & t_i \neq t_k \end{cases} \quad (3.249)$$

Note that when $t_i \neq t_k$, that we have nonoverlapping intervals since all of our intervals are disjoint per our construction. ■

3.5 Infinite-Dimensional Sampled-Data Kalman Filter

The first Kalman filter was derived for a discrete-time environment with finite-dimensional states by Kalman in 1960 [95]. One year later, Kalman and Bucy combined efforts to pose the Kalman-Bucy filter to treat continuous-time-measurement estimation problems [96]. Many physically motivated problems are set in a (more general) Hilbert space that is not necessarily of finite dimension. In 1967, Falb contributed the infinite-dimensional Kalman-Bucy filter (IKBF) [51]. Note that when we say “infinite-dimensional,” what we are really saying is that the states do not have to be finite length vectors: they can be functions or some other objects defined on a Hilbert space. The following table shows that with the addition of the new infinite-dimensional sampled-data Kalman filter (ISKF), linear filtering theory, consisting of the four filters of Table 3.1, forms a complete set of optimal estimation tools that can be applied in practice.

Before proceeding, we shall explicitly call attention to the two forms of the conditional error covariance that we previously defined in Equation (3.211) as we

	Discrete-time	Continuous-time
Finite-dimensional	Kalman filter (1960)	Kalman-Bucy filter (1961)
Infinite-dimensional	ISKF (2007)	IKBF (1967)

Table 3.1 Quartet of Kalman Filters

updated the estimate with a new observation and in Equation (3.228) in order to propagate the estimate from time t_i to time t_{i+1} .

Definition 89 (Conditional error covariance) *The propagated state estimator error is defined by*

$$\mathbf{e}(t_i^-) \triangleq \mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-) \quad (3.250)$$

and the zero-mean conditional error covariance operator for the propagated state estimator error is

$$\begin{aligned} \mathbf{P}(t_i^-) &\triangleq \Sigma[\mathbf{e}(t_i^-) | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] \\ &= E\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)] \diamond [\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)] | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\} \end{aligned} \quad (3.251)$$

where the realization of $\hat{\mathbf{x}}(t_i^-)$ is used in the second line since the measurement is given and the ordered sets $\mathbf{Z}(t_{i-1})$ and \mathbf{Z}_{i-1} represent the stochastic measurement history and measurement history sample, respectively, through time t_{i-1} as defined in Equations (3.204) and (3.205).

Similarly, the updated state estimator error is defined by

$$\mathbf{e}(t_i^+) \triangleq \mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+) \quad (3.252)$$

and the zero-mean conditional error covariance operator for the updated state estimator error is given as

$$\begin{aligned} \mathbf{P}(t_i^+) &\triangleq \Sigma[\mathbf{e}(t_i^+) | \mathbf{Z}(t_i) = \mathbf{Z}_i] \\ &= E\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)] \diamond [\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)] | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \end{aligned} \quad (3.253)$$

where the realization of $\hat{\mathbf{x}}(t_i^+)$ is used in the second line since the measurement is given and the ordered sets $\mathbf{Z}(t_i)$ and \mathbf{Z}_i represent the stochastic measurement history and measurement history sample, respectively, through time t_i as defined in Equations (3.204) and (3.205).

We pose the following lemma relating two conditional error covariances to their respective conditional state covariances. These relationships are common knowledge for finite-dimensional systems [129].

Lemma 90 *Given Definition 89, the following equivalences hold*

$$\Sigma[\mathbf{e}(t_i^-)|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] \equiv \Sigma[\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] \quad (3.254)$$

$$\Sigma[\mathbf{e}(t_i^+)|\mathbf{Z}(t_i) = \mathbf{Z}_i] \equiv \Sigma[\mathbf{x}(t_i)|\mathbf{Z}(t_i) = \mathbf{Z}_i] \quad (3.255)$$

Proof of Lemma 90 Per the definition of the error given in Equation (3.250), the left-hand side of Equation (3.254) is

$$\Sigma[\mathbf{e}(t_i^-)|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] = \Sigma[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] \quad (3.256)$$

Next, expand the right-hand side to get

$$\begin{aligned} & \Sigma[\mathbf{e}(t_i^-)|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] \\ &= E\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)] \diamond [\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)] | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\} \end{aligned} \quad (3.257)$$

$$= E\{[\mathbf{x}(t_i) - \mathcal{E}[\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})]] \diamond [\mathbf{x}(t_i) - \mathcal{E}[\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})]] | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\} \quad (3.258)$$

where lines one and two follow from the definition of the covariance and the definition of the conditional state estimator, respectively. Since the measurement history through time t_{i-1} is known, the conditional state estimator $\mathcal{E}[\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})]$ is really $E[\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}]$, which is just the conditional state estimate $\hat{\mathbf{x}}(t_i^-)$, thus we

get

$$\begin{aligned}\Sigma[\mathbf{e}(t_i^-)|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] \\ = E\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)] \diamond [\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)] | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\}\end{aligned}\quad (3.259)$$

$$= \Sigma[\mathbf{x}(t_i) | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] \quad (3.260)$$

since $\hat{\mathbf{x}}(t_i^-) = E[\mathbf{x}(t_i) | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}]$. A similar line of reasoning holds for the conditional error covariance following measurement update. ■

We are now ready to state in the form of a theorem the central result in this chapter: the infinite-dimensional sampled-data Kalman filter (ISKF).

Theorem 91 (ISKF) *Given: the measurement model of Definition 76 and the equivalent discrete-time dynamics model comprised of Definition 80. Thus, we have the following stochastic difference equations²³,*

$$\mathbf{x}(t_{i+1}) = \Phi(t_{i+1}, t_i) \mathbf{x}(t_i) + \mathbf{B}_d(t_i) \mathbf{u}(t_i) + \mathbf{G}_d(t_i) \mathbf{w}_d(t_i) \quad (3.261)$$

and

$$\mathbf{z}(t_i) = \mathbf{H}(t_i) \mathbf{x}(t_i) + \mathbf{v}(t_i) \quad (3.262)$$

where the dynamics model is further described by

$$\begin{array}{llll}\mathbf{x}(t_i) \in \widetilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X}) & \dots & \text{state vector} \\ \Phi(t_{i+1}, t_i) \in \mathcal{BLO}(\widetilde{\mathbb{X}}) & \dots & \text{state transition operator from time } t_i \text{ to } t_{i+1} \\ \mathbf{B}_d(t_i) \in \mathcal{LT}(\mathbb{U}, \widetilde{\mathbb{X}}) & \dots & \text{discrete-time input distributor transformation} \\ \mathbf{u}(t_i) \in \mathbb{U} & \dots & \text{known control input vector} \\ \mathbf{G}_d(t_i) \in \mathcal{BLT}(\widetilde{\mathbb{W}}, \widetilde{\mathbb{X}}) & \dots & \text{discrete-time noise distributor transformation} \\ \mathbf{w}_d(t_i) \in \widetilde{\mathbb{W}} = \mathbb{L}^2(\Omega, P; \mathbb{W}) & \dots & \text{zero-mean white Gaussian noise vector}\end{array}$$

and the components of the observation model are

²³ $\mathbf{G}_d(t_i)$ was not assumed to be the identity operator *here*, although without loss of generality, it can be.

$$\begin{aligned}
\mathbf{z}(t_i) &\in \tilde{\mathbb{Z}} = \mathbb{L}^2(\Omega, P; \mathbb{Z}) & \dots & \text{measurement vector} \\
\mathbf{H}(t_i) &\in \mathcal{BLT}(\tilde{\mathbb{X}}, \tilde{\mathbb{Z}}) & \dots & \text{measurement distributor transformation} \\
\mathbf{x}(t_i) &\in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X}) & \dots & \text{state vector} \\
\mathbf{v}(t_i) &\in \tilde{\mathbb{V}} = \mathbb{L}^2(\Omega, P; \mathbb{V}) & \dots & \text{measurement-corruption noise vector}
\end{aligned}$$

Additionally, we assume that $\mathbf{v}(t_j)$, $\mathbf{w}_d(t_i)$, and the initial state $\mathbf{x}(t_0)$ are mutually independent for all time. Thus, $\mathbf{x}(t_i)$ and $\mathbf{v}(t_j)$ are independent for all time.

The **ISKF algorithm** consists of a two-step recursive process following initial state and conditional error covariance estimates, which are actually the mean and covariance of the Gaussian random vector $\mathbf{x}(t_0)$:

$$\hat{\mathbf{x}}(t_0) = E[\mathbf{x}(t_0)] = \hat{\mathbf{x}}_0 \quad (3.263)$$

$$\mathbf{P}(t_0) = \mathbf{P}_0 \quad (3.264)$$

At time t_i , the filter-computed residual covariance and Kalman gain transformation are, respectively,

$$\mathbf{A}(t_i) \equiv \mathbf{H}(t_i) \mathbf{P}(t_i^-) \mathbf{H}^*(t_i) + \mathbf{R}(t_i) \quad (3.265)$$

$$\mathbf{K}(t_i) \equiv \mathbf{P}(t_i^-) \mathbf{H}^*(t_i) \mathbf{A}^{-1}(t_i) \in \mathcal{BLT}(\tilde{\mathbb{Z}}, \tilde{\mathbb{X}}) \quad (3.266)$$

When the conditional state estimator $\mathcal{E}[\mathbf{x}(t_i)|\mathbf{Z}(t_i)]$ is evaluated with the current measurement $\mathbf{z}(t_i) = \mathbf{z}_i$ it becomes a realization of $\hat{\mathbf{x}}(t_i^+)$, which we denote by $\hat{\mathbf{x}}(t_i^+)$, and thus

$$\begin{aligned}
\hat{\mathbf{x}}(t_i^+) &\triangleq E[\mathbf{x}(t_i)|\mathbf{Z}(t_i) = \mathbf{Z}_i] \\
&= \hat{\mathbf{x}}(t_i^-) + \mathbf{K}(t_i) \mathbf{r}(t_i)
\end{aligned} \quad (3.267)$$

where

$$\mathbf{r}(t_i) \equiv \mathbf{z}_i - \mathbf{H}(t_i) \hat{\mathbf{x}}(t_i^-) \quad (3.268)$$

is called the measurement residual. The corresponding conditional error covariance after the measurement update, defined in Equation (3.253), is given as

$$\mathbf{P}(t_i^+) = \mathbf{P}(t_i^-) - \mathbf{K}(t_i) \mathbf{H}(t_i) \mathbf{P}(t_i^-) \quad (3.269)$$

Next, the state estimator $\hat{\mathbf{x}}(t_i^+) \triangleq \mathcal{E}[\mathbf{x}(t_i) | \mathbf{Z}(t_i)]$ is propagated to time t_{i+1} using Equation (3.261) and becomes $\hat{\mathbf{x}}(t_{i+1}^-) \triangleq \mathcal{E}[\mathbf{x}(t_{i+1}) | \mathbf{Z}(t_i)]$, where it is then evaluated using the previous measurement history $\mathbf{Z}(t_i) = \mathbf{Z}_i$ to produce the realization $\hat{\mathbf{x}}(t_{i+1}^-)$, hence

$$\begin{aligned} \hat{\mathbf{x}}(t_{i+1}^-) &\triangleq E[\mathbf{x}(t_{i+1}) | \mathbf{Z}(t_i) = \mathbf{Z}_i] \\ &= \Phi(t_{i+1}, t_i) \hat{\mathbf{x}}(t_i^+) + \mathbf{B}_d(t_i) \mathbf{u}(t_i) \end{aligned} \quad (3.270)$$

The corresponding conditional error covariance, defined in Equation (3.251), is

$$\mathbf{P}(t_{i+1}^-) = \Phi(t_{i+1}, t_i) \mathbf{P}(t_i^+) \Phi^*(t_{i+1}, t_i) + \mathbf{G}_d(t_i) \mathbf{Q}_d(t_i) \mathbf{G}_d^*(t_i) \quad (3.271)$$

Note that t_i^- denotes the time just before incorporating the measurement taken at time t_i , i.e., it is the time to which the previous update is propagated, and time t_i^+ denotes the time at which the state is updated after the measurement was taken. Thus, the progression of time is: $t_0 < t_1^- < t_1 < t_1^+ < t_2^- < t_2 < t_2^+ \dots$.

The boxology for the ISKF is shown in Figure 3.13; it simply combines the previous boxologies for the stochastic LIMVUE seen in Figure 3.11, page 3-69, and the dynamics model boxology of Figure 3.12 on page 3-75.

Proof of Theorem 91 Equations (3.263) and (3.264) are initialization statements that do not need proving. Similarly, Equations (3.265), (3.266), and (3.268) are meaningful shorthand notations that are useful quantities to analyze during engineering studies of the problem at hand. Equations (3.270), (3.271), (3.267), and (3.269) remain to be proven.

We begin by substituting Equation (3.261) into the definition for the estimate $\hat{\mathbf{x}}(t_{i+1}^-)$ in Equation (3.270) and then simplifying

$$\hat{\mathbf{x}}(t_{i+1}^-) \triangleq E[\mathbf{x}(t_{i+1})|\mathbf{Z}(t_i) = \mathbf{Z}_i] \quad (3.272)$$

$$\begin{aligned} &= E\{[\Phi(t_{i+1}, t_i) \mathbf{x}(t_i) + \mathbf{B}_d(t_i) \mathbf{u}(t_i) \\ &\quad + \mathbf{G}_d(t_i) \mathbf{w}_d(t_i)]|\mathbf{Z}(t_i) = \mathbf{Z}_i\} \end{aligned} \quad (3.273)$$

Then,

$$\begin{aligned} \hat{\mathbf{x}}(t_{i+1}^-) &= \Phi(t_{i+1}, t_i) E[\mathbf{x}(t_i)|\mathbf{Z}(t_i) = \mathbf{Z}_i] + \mathbf{B}_d(t_i) \mathbf{u}(t_i) \\ &\quad + \mathbf{G}_d(t_i) E[\mathbf{w}_d(t_i)|\mathbf{Z}(t_i) = \mathbf{Z}_i] \end{aligned} \quad (3.274)$$

$$= \Phi(t_{i+1}, t_i) \hat{\mathbf{x}}(t_i^+) + \mathbf{B}_d(t_i) \mathbf{u}(t_i) \quad (3.275)$$

where the second line follows from the definition of $\hat{\mathbf{x}}(t_i^+)$ and the fact that the dynamics noise is (assumed to be) zero-mean and independent of \mathbf{v} and hence of $\mathbf{Z}(t_i)$ as well. Thus Equation (3.270) results as proposed.

Next, we use the propagation Equation (3.261), for $\hat{\mathbf{x}}(t_{i+1}^-)$ in the equivalence of the conditional error covariance to the conditional state covariance given in Equation (3.254) to obtain

$$\mathbf{P}(t_{i+1}^-) = \Sigma[\mathbf{x}(t_{i+1})|\mathbf{Z}(t_i) = \mathbf{Z}_i] \quad (3.276)$$

$$= \Sigma[\Phi(t_{i+1}, t_i) \mathbf{x}(t_i) + \mathbf{B}_d(t_i) \mathbf{u}(t_i) + \mathbf{G}_d(t_i) \mathbf{w}_d(t_i)|\mathbf{Z}(t_i) = \mathbf{Z}_i] \quad (3.277)$$

Dropping the $\mathbf{B}_d(t_i) \mathbf{u}(t_i)$ term as it is known and thus does not contribute to the covariance and then expanding yields

$$\mathbf{P}(t_{i+1}^-) = \Sigma[\Phi(t_{i+1}, t_i) \mathbf{x}(t_i) + \mathbf{G}_d(t_i) \mathbf{w}_d(t_i)|\mathbf{Z}(t_i) = \mathbf{Z}_i] \quad (3.278)$$

Expanding again,

$$\begin{aligned}
\mathbf{P}(t_{i+1}^-) &= \Sigma[\Phi(t_{i+1}, t_i) \mathbf{x}(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i] \\
&\quad + \Sigma\{\Phi(t_{i+1}, t_i) \mathbf{x}(t_i), \mathbf{G}_d(t_i) \mathbf{w}_d(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \\
&\quad + \Sigma\{\mathbf{G}_d(t_i) \mathbf{w}_d(t_i), \Phi(t_{i+1}, t_i) \mathbf{x}(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \\
&\quad + \Sigma[\mathbf{G}_d(t_i) \mathbf{w}_d(t_i)]
\end{aligned} \tag{3.279}$$

where the conditioning for the fourth term was dropped since \mathbf{w}_d and $\mathbf{Z}(t_{i-1})$ are independent as previously noted. The first term in Equation (3.279) is rewritten in expectation notation as

$$\begin{aligned}
&\Sigma[\Phi(t_{i+1}, t_i) \mathbf{x}(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i] \\
&= E\{\Phi(t_{i+1}, t_i) [\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)] \diamond \Phi(t_{i+1}, t_i) [\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)] | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \tag{3.280}
\end{aligned}$$

$$= \Phi(t_{i+1}, t_i) E\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)] \diamond [\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)] | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \Phi^*(t_{i+1}, t_i) \tag{3.281}$$

$$= \Phi(t_{i+1}, t_i) \Sigma[\mathbf{x}(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i] \Phi^*(t_{i+1}, t_i) \tag{3.282}$$

$$= \Phi(t_{i+1}, t_i) \mathbf{P}(t_i^+) \Phi^*(t_{i+1}, t_i) \tag{3.283}$$

where the second equality employs Lemma 43, from page 3-26, and lines three and four follow from definitions for a covariance operator and then the conditional error covariance operator. The second and third terms of Equation (3.279) are cross-covariance terms for independent random vectors, with at least one being zero-mean, and are thus zero. The fourth term of Equation (3.279) is expanded using the expectation notation, while noting that \mathbf{w}_d is a zero-mean stochastic noise process, to obtain

$$\Sigma[\mathbf{G}_d(t_i) \mathbf{w}_d(t_i)] = E[\mathbf{G}_d(t_i) \mathbf{w}_d(t_i) \diamond \mathbf{G}_d(t_i) \mathbf{w}_d(t_i)] \tag{3.284}$$

$$= \mathbf{G}_d(t_i) E[\mathbf{w}_d(t_i) \diamond \mathbf{w}_d(t_i)] \mathbf{G}_d^*(t_i) \tag{3.285}$$

$$= \mathbf{G}_d(t_i) \mathbf{Q}_d(t_i) \mathbf{G}_d^*(t_i) \tag{3.286}$$

where Lemma 43 is used to produce the second equality and then \mathbf{Q}_d is simply the covariance of \mathbf{w}_d defined in Equation (3.225). Therefore, with each of the terms systematically addressed, Equation (3.279) becomes Equation (3.271).

The updated state estimator, $\hat{\mathbf{x}}(t_i^+)$, in Equation (3.267) and the conditional error covariance, $\mathbf{P}(t_i^+)$, in Equation (3.269), are estimators for the state and the conditional error covariance for the stochastic LIMVUE given in Theorem 78, as noted in the remark following the theorem, for the i th measurement $\mathbf{Z}(t_i) = \mathbf{Z}_i$. ■

As reported earlier, the state and measurement-corruption noise were assumed to be independent processes. The following lemma shows why this is true for the discrete-time case.

Lemma 92 *The state, $\mathbf{x}(t_i)$, and measurement-corruption noise, $\mathbf{v}(t_j)$, are independent for all time $t_i, t_j \in \mathbb{T}$.*

Proof of Lemma 92²⁴ Recall Equation (3.224) from the definition of the dynamics model (where we have decremented all of the time indices):

$$\mathbf{x}(t_i) = \Phi(t_i, t_{i-1}) \mathbf{x}(t_{i-1}) + \mathbf{B}_d(t_{i-1}) \mathbf{u}(t_{i-1}) + \mathbf{G}_d(t_{i-1}) \mathbf{w}_d(t_{i-1}) \quad (3.287)$$

Using this equation, we substitute in for $\mathbf{x}(t_{i-1})$ and we get

$$\begin{aligned} \mathbf{x}(t_i) = & \Phi(t_i, t_{i-1}) [\Phi(t_{i-1}, t_{i-2}) \mathbf{x}(t_{i-2}) + \mathbf{B}_d(t_{i-2}) \mathbf{u}(t_{i-2}) \\ & + \mathbf{G}_d(t_{i-2}) \mathbf{w}_d(t_{i-2})] + \mathbf{B}_d(t_{i-1}) \mathbf{u}(t_{i-1}) + \mathbf{G}_d(t_{i-1}) \mathbf{w}_d(t_{i-1}) \end{aligned} \quad (3.288)$$

which can be written as

$$\begin{aligned} \mathbf{x}(t_i) = & \Phi(t_i, t_{i-2}) \mathbf{x}(t_{i-2}) + \Phi(t_i, t_{i-1}) [\mathbf{B}_d(t_{i-2}) \mathbf{u}(t_{i-2}) \\ & + \mathbf{G}_d(t_{i-2}) \mathbf{w}_d(t_{i-2})] + \mathbf{B}_d(t_{i-1}) \mathbf{u}(t_{i-1}) + \mathbf{G}_d(t_{i-1}) \mathbf{w}_d(t_{i-1}) \end{aligned} \quad (3.289)$$

²⁴The proof of this lemma closely follows Maybeck [129] for finite-dimensional random vectors.

Now substitute in for $\mathbf{x}(t_{i-2})$ to continue the pattern

$$\begin{aligned}\mathbf{x}(t_i) = & \Phi(t_i, t_{i-2})[\Phi(t_{i-2}, t_{i-3})\mathbf{x}(t_{i-3}) + \mathbf{B}_d(t_{i-3})\mathbf{u}(t_{i-3}) \\ & + \mathbf{G}_d(t_{i-3})\mathbf{w}_d(t_{i-3})] + \Phi(t_i, t_{i-1})[\mathbf{B}_d(t_{i-2})\mathbf{u}(t_{i-2}) \\ & + \mathbf{G}_d(t_{i-2})\mathbf{w}_d(t_{i-2})] + \mathbf{B}_d(t_{i-1})\mathbf{u}(t_{i-1}) + \mathbf{G}_d(t_{i-1})\mathbf{w}_d(t_{i-1})\end{aligned}\quad (3.290)$$

Simplifying,

$$\begin{aligned}\mathbf{x}(t_i) = & \Phi(t_i, t_{i-3})\mathbf{x}(t_{i-3}) \\ & + \Phi(t_i, t_{i-2})[\mathbf{B}_d(t_{i-3})\mathbf{u}(t_{i-3}) + \mathbf{G}_d(t_{i-3})\mathbf{w}_d(t_{i-3})] \\ & + \Phi(t_i, t_{i-1})[\mathbf{B}_d(t_{i-2})\mathbf{u}(t_{i-2}) + \mathbf{G}_d(t_{i-2})\mathbf{w}_d(t_{i-2})] \\ & + \mathbf{B}_d(t_{i-1})\mathbf{u}(t_{i-1}) + \mathbf{G}_d(t_{i-1})\mathbf{w}_d(t_{i-1})\end{aligned}\quad (3.291)$$

The nested pattern is now clear, and thus we write Equation (3.291) as

$$\mathbf{x}(t_i) = \Phi(t_i, t_0)\mathbf{x}(t_0) + \sum_{k=1}^i \Phi(t_i, t_k)[\mathbf{B}_d(t_{k-1})\mathbf{u}(t_{k-1}) + \mathbf{G}_d(t_{k-1})\mathbf{w}_d(t_{k-1})]\quad (3.292)$$

where we used the fact that $\Phi(t_i, t_i)$, for any time t_i , is equivalent to the identity operator. Since $\mathbf{v}(t_j)$ is independent of each of the terms in Equation (3.292), $\mathbf{x}(t_i)$ and $\mathbf{v}(t_j)$ are mutually independent random vectors for all time $t_i, t_j \in \mathbb{T}$. ■

3.6 Generalized Infinite-Dimensional Multiple Model Adaptive Estimation

The system description entails a detailed accounting of all of the parameters used in the structure of the models and the statistics describing the dynamics and measurement noises. Specifically, the system is determined by knowledge of the true values for Φ , \mathbf{B}_d , \mathbf{G}_d , \mathbf{H} , \mathbf{Q}_d , \mathbf{R} , $\hat{\mathbf{x}}_0$, and \mathbf{P}_0 . When a subset of the model parameters are uncertain, we can characterize this subset of uncertain parameters as stochastic processes. For this research, we restrict ourselves to a subset of these parameters,

and since their values are uncertain (but assumed constant), we express them in terms of the components of a vector random-constant stochastic process $\mathbf{a}(\cdot, \cdot)$. The random-constant stochastic process, indexed by the times \mathbb{T} , is a constant random vector for all times $t_i \in \mathbb{T}$, i.e., $\mathbf{a}(t_i, \cdot) \equiv \mathbf{a}(t_i) \in \tilde{\mathbb{A}}$ and for a given $\omega \in \Omega$, the realization is $\mathbf{a}(t_i, \omega) \equiv \mathbf{a} \in \mathbb{A}$, which is independent of the time index since it is assumed to be a constant for all time. More general stochastic process models than random-constant processes can allow the parameter to be time-varying, and this can give rise to an interactive multiple model (IMM) rather than an MMAE algorithm, as discussed in Section 2.4.7, page 2-38. Section 2.3.3.1 has further information for the finite-dimensional case.

Note that the form of the equations for the elemental filters and the state and parameter estimates look exactly the same as those equations appearing in Sections 2.3.3.3 in 2.3.4. However, they are not strictly the same since in this chapter we are dealing with the more general case of vectors in a Hilbert space, i.e., they may be infinite-dimensional vectors, and the matrices for the finite-dimensional case are now, in general, allowed to be transformations — thus, we have the generalized infinite-dimensional multiple model adaptive estimation (GIMMAE).

3.6.1 Elemental Filters. Each elemental filter in the bank is based upon a different hypothesis for the parameter values used to model the real world system, i.e., the k th elemental filter design model is constructed assuming that $\mathbf{a}(t_i) = \mathbf{a}_k$. The discrete-time model equations for the k th elemental filter are

$$\mathbf{x}_k(t_{i+1}) = \Phi_k(t_{i+1}, t_i) \mathbf{x}_k(t_i) + \mathbf{B}_{dk}(t_i) \mathbf{u}(t_i) + \mathbf{G}_{dk}(t_i) \mathbf{w}_{dk}(t_i) \quad (3.293)$$

$$\mathbf{z}(t_i) = \mathbf{H}_k(t_i) \mathbf{x}_k(t_i) + \mathbf{v}_k(t_i) \quad (3.294)$$

where the properties of $\Phi_k(t_{i+1}, t_i)$, $\mathbf{B}_{dk}(t_i)$, $\mathbf{G}_{dk}(t_i)$, $\mathbf{H}_k(t_i)$, $\mathbf{Q}_{dk}(t_i)$, and $\mathbf{R}_k(t_i)$ were discussed in the previous sections.

The correctness or validity of each hypothesis, $\mathbf{a}(t_i) = \mathbf{a}_k$, is ordinarily obtained through an analysis of the filter residuals, the difference between the ob-

served measurement and the predicted measurement, $\mathbf{r}_k(t_i) = \mathbf{z}_i - \mathbf{H}_k(t_i) \hat{\mathbf{x}}_k(t_i^-)$ [129]. This “correctness” information is also coded in the hypothesis conditional probability $p_k(t_i)$, which is defined as the probability that $\mathbf{a}(t_i)$ assumes the value \mathbf{a}_k (for $k = 1, 2, \dots, K$), conditioned on the observed measurement history to time t_i [130, 132]

$$p_k(t_i) = \text{pr}\{\mathbf{a}(t_i) = \mathbf{a}_k | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \quad (3.295)$$

such that

$$p_k(t_i) \geq 0 \text{ for all } k \quad \text{and} \quad \sum_{k=1}^K p_k(t_i) = 1 \quad (3.296)$$

A close inspection of Equations (2.44) through (2.46), specifically, Equation (2.45), on page 2-30, shows that the PDF for the Gaussian distributed random vector does not exist on a general Hilbert space since letting measurement dimension m tend to infinity results in β equal to zero; hence we would have $f = 0$, which is not a proper PDF²⁵. However, it is interesting to note that the hypothesis conditional probabilities calculated using Equation (2.43), found on page 2-29, are independent of m , since it factors out of both the numerator and the sum of K terms in the denominator. So, assuming that the initial probability $p_k(t_0)$ for all k is known or well modeled, for example, as $p_k(t_0) = 1/K$ for $k = 1, \dots, K$, hypothesis conditional probabilities are determined as

$$p_k(t_i) = \frac{\hat{f}_{\mathbf{z}(t_i)|\mathbf{a}(t_i), \mathbf{Z}(t_{i-1})}(\mathbf{z}_i | \mathbf{a}_k, \mathbf{Z}_{i-1}) p_k(t_{i-1})}{\sum_{j=1}^K \hat{f}_{\mathbf{z}(t_i)|\mathbf{a}(t_i), \mathbf{Z}(t_{i-1})}(\mathbf{z}_i | \mathbf{a}_j, \mathbf{Z}_{i-1}) p_j(t_{i-1})} \quad (3.297)$$

This is an iteration expressed in terms of the previous values $p_k(t_{i-1})$, where the scaled conditional probability “density” function, as denoted by \hat{f} , is a Gaussian-like function with “mean” $\mathbf{H}_k(t_i) \hat{\mathbf{x}}_k(t_i^-)$ and “covariance” (operator) $\mathbf{A}_k(t_i)$

$$\hat{f}_{\mathbf{z}(t_i)|\mathbf{a}(t_i), \mathbf{Z}(t_{i-1})}(\mathbf{z}_i | \mathbf{a}_k, \mathbf{Z}_{i-1}) = \hat{\beta}_k(t_i) \exp \left\{ -\frac{1}{2} L_k(t_i) \right\} \quad (3.298)$$

²⁵We defined the Gaussian-distributed random vector using the characteristic equation, see page 3-32, because the PDF is ill-defined for the infinite-dimensional case.

where the modified scale factor is now

$$\hat{\beta}_k(t_i) = \frac{1}{\|\mathbf{A}_k(t_i)\|^{1/2}} \quad (3.299)$$

and where the likelihood quotient in Equation (3.298), which is a measure of the “correctness” of the parameter values for this particular model [130], is the weighted inner product

$$L_k(t_i) = \langle \mathbf{r}_k(t_i), \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i) \rangle \quad (3.300)$$

where $\mathbf{r}_k(t_i) = \mathbf{z}_i - \mathbf{H}_k(t_i) \hat{\mathbf{x}}_k(t_i^-)$ and $\mathbf{A}_k(t_i)$ are the residual and associated residual covariance calculated by the k th Kalman filter as in Equations (3.268) and (3.265), respectively. Note that using \hat{f} (which is not a true PDF since the “volume” under the \hat{f} function “surface” is *not* unity) rather than f (which is ill-defined in the infinite-dimensional case) has no impact on the operation of the MMAE. Furthermore, the denominator in Equation (3.297) is simply the sum of all K numerators, and it is thus the appropriate scale factor to guarantee that the $p_k(t_i)$ values so generated will always sum to one. Finally, note that $\hat{\beta}_k(t_i)$ is just a scale factor and that the most important information to be retrieved from this “density” function is contained in $L_k(t_i)$; hence the fact that we do not have true PDFs in Equation (3.297) in the strict sense is not problematic.

It has been shown, for the finite-dimensional case [94, 129], that the sequence of residuals $\{\mathbf{r}_k(t_i)\}$ resulting from linear filtering forms a zero-mean white Gaussian sequence with known residual covariance $\mathbf{A}_k(t_i)$. Thus, if a filter model matches the “true” system, then the residual $\mathbf{r}_k(t_i)$ should be a zero-mean white Gaussian process with known residual covariance $\mathbf{A}_k(t_i)$.

3.6.2 State and Parameter Estimates. This estimation technique uses the information in all of the Kalman filter residuals to estimate the “true” parameter vector in effect. This technique is optimal when there is a unique filter for each pos-

sible combination of parameter values, which is only possible when the parameter(s) of interest takes on just a finite number of possible values. We shall populate the filter bank with K filters, each based on one of the K unique parameter vectors: $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_K\}$.

From the Bayesian point of view, the MMAE framework can be used to compute a state (or parameter) estimate that is characterized by minimizing the MSE between the predicted and measured state estimates; this is most often called an MMSE estimate and is the conditional mean. We identify the Bayesian estimate as the standard MMAE estimate and write it as [130, 132]

$$\hat{\mathbf{x}}_{\text{MMAE}}(t_i^+) \triangleq E\{\mathbf{x}(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i\} = \sum_{k=1}^K \hat{\mathbf{x}}_k(t_i^+) p_k(t_i) \quad (3.301)$$

where $\hat{\mathbf{x}}_k(t_i^+)$ is the state estimate generated by the k th Kalman filter based on the assumption that the parameter vector $\mathbf{a}(t_i) = \mathbf{a}_k$ for all t_i . The conditional covariance of $\mathbf{x}(t_i)$ computed by the MMAE is given by Equation (2.50) for finite-dimensional systems [130] and for infinite-dimensional systems is

$$\begin{aligned} \mathbf{P}_{\text{MMAE}}(t_i^+) &\triangleq E\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}_{\text{MMAE}}(t_i^+)] \diamond [\mathbf{x}(t_i) - \hat{\mathbf{x}}_{\text{MMAE}}(t_i^+)] | \mathbf{Z}(t_i) = \mathbf{Z}_i\} \end{aligned} \quad (3.302)$$

$$= \sum_{k=1}^K \{\mathbf{P}_k(t_i^+) + [\hat{\mathbf{x}}_k(t_i^+) - \hat{\mathbf{x}}_{\text{MMAE}}(t_i^+)] \diamond [\hat{\mathbf{x}}_k(t_i^+) - \hat{\mathbf{x}}_{\text{MMAE}}(t_i^+)]\} p_k(t_i) \quad (3.303)$$

where $\mathbf{P}_k(t_i^+)$ is the state error covariance computed by the k th Kalman filter. Similarly, the parameter estimate is given by

$$\hat{\mathbf{a}}_{\text{MMAE}}(t_i^+) \triangleq E\{\mathbf{a}(t_i) | \mathbf{Z}(t_i) = \mathbf{Z}_i\} = \sum_{k=1}^K \mathbf{a}_k p_k(t_i) \quad (3.304)$$

with conditional covariance of $\mathbf{a}(t_i)$ for finite-dimensional systems [129] as given in Equation (2.54), and adapted for infinite-dimensional systems as

$$\mathbf{P}_{\mathbf{a},\text{MMAE}}(t_i^+) \triangleq E \left\{ [\mathbf{a}(t_i) - \hat{\mathbf{a}}_{\text{MMAE}}(t_i^+)] \diamond [\mathbf{a}(t_i) - \hat{\mathbf{a}}_{\text{MMAE}}(t_i^+)] | \mathbf{Z}(t_i) = \mathbf{Z}_i \right\} \quad (3.305)$$

$$= \sum_{k=1}^K [\mathbf{a}_k - \hat{\mathbf{a}}_{\text{MMAE}}(t_i^+)] \diamond [\mathbf{a}_k - \hat{\mathbf{a}}_{\text{MMAE}}(t_i^+)] p_k(t_i) \quad (3.306)$$

3.7 Summary

The early part of this chapter focused on defining various mathematical concepts needed to construct the linear infinite-dimensional minimum variance unbiased estimator (LIMVUE) rigorously — this is a central tool in the building of the infinite-dimensional sampled-data Kalman filter (ISKF). Along the way, we introduced the illustrative boxology technique that we use to convey the ISKF development, from defining the primary probability space to the spaces occupied by the random vectors, to the probability spaces induced by the random vectors representing the noises, the observations, and the state. We generalized the LIMVUE for stochastic processes to create a stochastic sequential estimator. Since the majority of the problems that we study are described by infinite-dimensional continuous-time models, we extended the known finite-dimensional method for creating an equivalent discrete-time model from its corresponding continuous-time model for infinite-dimensional models. The dynamics model provides us with a tool to propagate the Gaussian conditional state “density” between measurements, the first two moments of which are estimated optimally using the stochastic LIMVUE. Then, we assembled the pieces to form the ISKF, thus completing the array of filtering techniques that began with Kalman’s first (discrete-time) filter [95] and the continuous-time Kalman-Bucy filter [96] shortly thereafter; subsequently, Falb’s [51] extension of the Kalman-Bucy filter (for continuous-time measurements) to encompass systems with an infinite-dimensional continuous-time description by developing it on a Hilbert space. The chapter closed with a short discussion and report on the modified formulae for the multiple model adaptive estimation (MMAE) methodology as applicable to a bank of elemental ISKFs — the GIMMAE.

IV. An Example: The Stochastic Heat Equation

4.1 Introduction

The state space model used in this stochastic estimation research is based on a pair of mathematical expressions: a state equation that defines the evolution of a dynamic process through time and a measurement equation that defines the observation process. In general, these equations may be either stochastic or deterministic; we shall investigate the stochastic case. The state space model may be based on differential or difference equations; our work encompasses both varieties. When presented with a continuous-time model featuring differential equations we shall re-express them using their equivalent discrete-time difference equations. Finally, while the dimension of the state space model is allowed to be infinite for theoretical purposes, for computational purposes, we must have finite-dimensional equations. Therefore, we shall present a straight-forward, yet novel, method for reposing the problem on a finite-dimensional subspace.

In this chapter we will apply the theoretical methods developed in the preceding chapter to a physically meaningful problem. We will estimate the temperature along a slender cylindrical rod modeled by the stochastic heat equation, a parabolic partial differential equation (PDE), using noise-corrupted finite-dimensional measurements¹. This example is a special case of the general theory developed in Chapter III since our state, the temperature, is a scalar, while the observations are recorded in a finite-dimensional vector of scalars. Many textbooks on PDEs, such as Berg and McGregor [20] written for mathematicians, Farlow [52] for scientists and engineers, and Gockenbach [62] for numerical-computational scientists and engineers, contain an exposition on the heat equation. However, all of these texts solve the deterministic problem as if the model were exact. Phillipson [161] also solved the deterministic

¹This example was inspired by Example 5.39 in Curtain and Pritchard [38].

problem, but he employed a least squares approach that admitted that there was some uncertainty in the temperature (or state of the system). He employed the Galerkin method² to find two approximate solutions for the state expressed as a linear combination of the eigenfunctions of the system and another useful ad hoc approach using a linear combination of cubic splines to speed up convergence [161].

In a recent paper, Leland [116] presented a method for treating parabolic PDEs (such as the stochastic heat equation problem) using a maximum likelihood estimator for the parameter of interest, such as the thermal diffusivity. He used an approximate time-invariant one-step predictor to avoid solving state and covariance equations; additionally, his system was presumed to be in steady state. This paper by Leland is representative of the other papers reviewed during this research in that it does not include a measurement model to take into account the measurement process.

In this research, we employ an evolution equation to model the time-varying dynamics of the system in question *and* a measurement model in order to estimate the state of the system optimally. As we shall soon demonstrate, the method that we employ to create the essentially-equivalent finite-dimensional discrete-time model from the infinite-dimensional continuous-time model allows us to use the infinite-dimensional sampled-data Kalman filter (ISKF) without additional approximations. The resulting algorithm looks and behaves like the finite-dimensional sampled-data Kalman filter that was reported in Chapter II.

4.2 Mathematical System Model

Creation of the mathematical system model is the first step in model-based estimation. We have used familiar models for the dynamics and measurement processes for two primary reasons: (1) to emphasize the applicability of the theory developed in the previous chapter and (2) to illustrate the techniques developed in this

²Numerous texts contain a paragraph, section, or chapter devoted to the explication of the Galerkin method used to solve PDEs [89, 30, 62].

chapter to transform an infinite-dimensional problem into an essentially-equivalent finite-dimensional problem that one can easily implement on a digital computer.

4.2.1 Preliminary Background: The Heat Equation. The purpose of this extended example is to demonstrate an exact method for employing the ISKF to estimate the temperature profile along the length of a slender cylindrical rod over time. Let's begin by describing the deterministic heat equation with Neumann boundary conditions in one dimension augmented by a variable heat source [52]:

$$\frac{d}{dt}x(t, \rho) = \kappa \frac{\partial^2}{\partial \rho^2}x(t, \rho) + u(t, \rho), \quad 0 < \rho < 1, \quad 0 < t < \infty \quad (4.1)$$

$$\frac{\partial}{\partial \rho}x(t, 0) = 0, \quad 0 < t < \infty \quad (4.2)$$

$$\frac{\partial}{\partial \rho}x(t, 1) = 0, \quad 0 < t < \infty \quad (4.3)$$

$$x(0, \rho) = x_0(\rho), \quad 0 < \rho < 1 \quad (4.4)$$

where x is the temperature in degrees Celsius ($^{\circ}\text{C}$), t is the time in seconds, $\kappa > 0$ is the thermal diffusivity constant of the material in square meters per second, the heat source is at u $^{\circ}\text{C}$, and the position along the rod is indicated by ρ . Additionally, the rod has an initial temperature of $x_0(\rho)$ and is laterally-insulated with insulated ends (boundaries). Now let's add a random component to the heat source — this also makes the temperature random — then Equations (4.1) through (4.4) become

$$\frac{d}{dt}\mathbf{x}(t, \rho) = \kappa \frac{\partial^2}{\partial \rho^2}\mathbf{x}(t, \rho) + u(t, \rho) + \mathbf{w}(t, \rho), \quad 0 < \rho < 1, \quad 0 < t < \infty \quad (4.5)$$

$$\frac{\partial}{\partial \rho}\mathbf{x}(t, 0) = 0, \quad 0 < t < \infty \quad (4.6)$$

$$\frac{\partial}{\partial \rho}\mathbf{x}(t, 1) = 0, \quad 0 < t < \infty \quad (4.7)$$

$$\mathbf{x}(0, \rho) = \mathbf{x}_0(\rho), \quad 0 < \rho < 1 \quad (4.8)$$

where \mathbf{w} is a zero-mean white Gaussian noise process with strength Q and \mathbf{x}_0 is a

Gaussian random variable with mean \hat{x}_0 and covariance P_0 . Note that Equations (4.5) through (4.8) only represent the stochastic heat equation in a formal sense since the solution is not properly defined in this “white-noise” continuous-time notation. A properly defined model was given in Section 3.4 and we will use those results in our example to describe properly the mathematical system model used for estimating the temperature (state) of the slender cylindrical rod with noise-corrupted measurements.

4.2.2 Model Mapping. In Chapter I we introduced a concept for mapping an infinite-dimensional continuous-time model, a projection of the real world onto a linear systems mind-set, to the equivalent infinite-dimensional discrete-time model; the infinite-dimensional sampled-data Kalman filter (ISKF) described in Section 3.5 was derived specifically for this case. We continued by projecting this equivalent model onto a finite-dimensional subspace to produce an *essentially-equivalent* finite-dimensional discrete-time model that we can use to design an optimal filter with which to estimate the state — by means of the sampled-data Kalman filter. Figure 4.1, an important part of Figure 1.1, gives an overview of the process. In *this* section, we map the infinite-dimensional continuous-time model to the equivalent infinite-dimensional discrete-time model, i.e., we execute the optimal discretization of the time variable, the conceptual \mathcal{T}_{opt} operation, using the technique and results presented in Section 3.4. In the next two sections we prepare the model for the Kalman filter by projecting our equivalent infinite-dimensional discrete-time model onto a finite-dimensional subspace; with this conceptual \mathcal{S}_{opt} operator we obtain the essentially-equivalent finite-dimensional discrete-time model. Then we construct the conceptual \mathcal{F}_{opt} by determining the finite-dimensional components of the sampled-data Kalman filter in Section 4.3. In Chapter V we use the results from this chapter to simulate the flow of heat through the slender cylindrical rod and to estimate the temperature using noisy measurements.

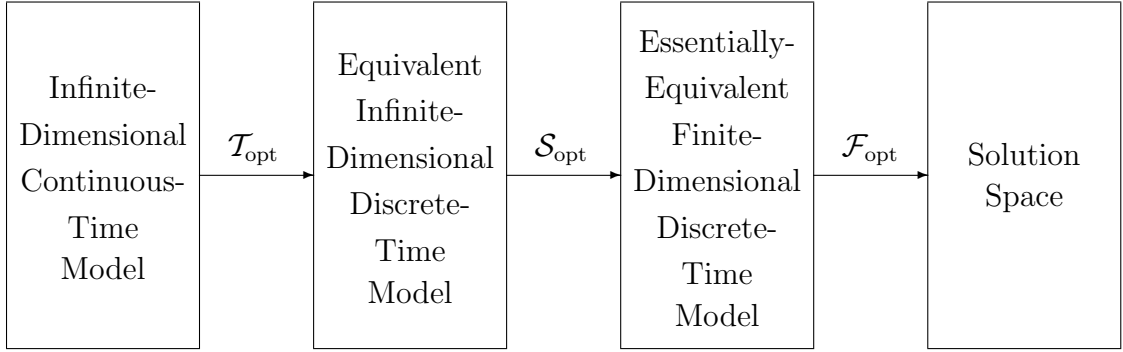


Figure 4.1 Mapping the Infinite-Dimensional Continuous-Time Model to an Essentially-Equivalent Finite-Dimensional Discrete-Time Model.

4.2.3 Discrete-Time Measurement Model. Let the following integral represent the operation of a sensor collecting information about the temperature of the rod at time t_i

$$\mathbf{z}(t_i) = \int_0^1 \mathbf{x}(t_i, \rho) d\rho + \mathbf{v}(t_i) \quad (4.9)$$

where \mathbf{z} , \mathbf{x} , and \mathbf{v} are (scalar) stochastic functions in Hilbert spaces of random variables. However, this model would only allow us calculate the average temperature of the slender cylindrical rod. So we propose to partition the integral into M segments to represent a set of M sensors positioned along M equi-length sections of the rod

$$\int_0^1 \mathbf{x}(t_i, \rho) d\rho = \int_0^{1/M} \mathbf{x}(t_i, \rho) d\rho + \cdots + \int_{(M-1)/M}^1 \mathbf{x}(t_i, \rho) d\rho \quad (4.10)$$

If we were to evaluate Equation (4.10) directly, then we would lose all of the spatial information. So, in order to preserve the spatial information, we propose to stack the M integrals in an observation vector to represent the contributions of M individual sensors

$$\mathbf{z}(t_i) = \begin{bmatrix} \int_0^{1/M} \mathbf{x}(t_i, \rho) d\rho \\ \vdots \\ \int_{(M-1)/M}^1 \mathbf{x}(t_i, \rho) d\rho \end{bmatrix} + \mathbf{v}(t_i) = \mathbf{H} \mathbf{x}(t_i) + \mathbf{v}(t_i) \quad (4.11)$$

where $\mathbf{z}(t_i)$ is a random M -vector of observations at time t_i of the temperature $\mathbf{x}(t_i, \rho)$ as a function of position ρ , \mathbf{H} is a linear transformation defined by the vector of M integrals acting on the state, and $\mathbf{v}(t_i)$ is a random noise vector. More specifically, the random vector $\mathbf{z}(t_i) \in \tilde{\mathbb{Z}} = \mathbb{L}^2(\Omega, P; \mathbb{Z})$ is a Lebesgue \mathbb{L}^2 function³ (in a separable Hilbert space of functions) that maps the sample space⁴ Ω , to the realization space $\mathbb{Z} = \mathbb{R}^M$, as shown in Figure 4.2, while the scalar random variable $\mathbf{x}(t_i) \in \tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{X})$, which is also a Lebesgue \mathbb{L}^2 function, maps the sample space to another separable Hilbert space of Lebesgue \mathbb{L}^2 functions on the interval $0 \leq \rho \leq 1$ written as: $\mathbb{X} = \mathbb{L}^2([0, 1], \mathbb{R})$. We sometimes substitute the shorthand $\mathbb{L}_{[0,1]}^2$ for the more explicit notation $\mathbb{L}^2([0, 1], \mathbb{R})$. A realization of the observation at time t_i is labeled $\mathbf{z}_i \in \mathbb{R}^M$. The zero-mean Gaussian noise process \mathbf{v} has covariance matrix $\mathbf{R}(t_i)$ at time t_i and covariance kernel

$$E [\mathbf{v}(t_i) \mathbf{v}^T(t_j)] = \begin{cases} \mathbf{R}(t_i), & t_i = t_j \\ \mathbf{0}, & t_i \neq t_j \end{cases} \quad (4.12)$$

with $\mathbf{R}(t_i), \mathbf{0} \in \mathbb{R}^{M \times M}$, thus \mathbf{v} is also a white process.

4.2.4 Continuous-Time Dynamics Model. The temperature distribution of a slender cylindrical rod is well modeled by a scalar heat equation with additive noise. Thus, we begin by writing the scalar heat equation as a stochastic differential equation in differential form (versus the familiar derivative form) to guarantee the

³Recall from Chapter III that the Lebesgue \mathbb{L}^2 functions form a Banach space with finite norm $\|x(t_i)\| = \left[\int_0^1 |x(t_i, \rho)|^2 d\rho \right]^{1/2}$ as defined in an example on page 3-7. Furthermore, the \mathbb{L}^2 functions are absolutely square integrable. When associated with an inner product, the \mathbb{L}^2 Lebesgue functions form a Hilbert space [122]. We need the completeness of a Hilbert space to assure ourselves that all of its Cauchy sequences converge to a limit in the space and thus any sequence of random vectors will also converge to a limit within the space [154]. An example on page 3-27 gives the interpretation of the norm for Lebesgue functions representing the mapping induced by the random vectors.

⁴Definition 41, on page 3-24, defines the complete probability space, denoted by the triplet (Ω, \mathcal{F}, P) , that we employ in this research, where Ω is a non-empty set called the sample space, \mathcal{F} is a σ -field which consists of all of the subsets of Ω , called events, and P is the probability measure.

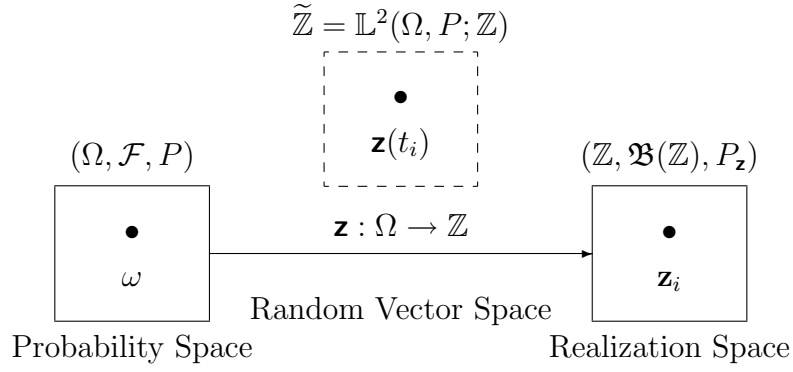


Figure 4.2 Boxology of the Random Measurement Vector.

existence and uniqueness of the solution [38], just as we did in Equation (3.223) on page 3-72

$$\begin{aligned} d\mathbf{x}(t, \rho) &= [F(\rho) \mathbf{x}(t, \rho) + B u(t, \rho)] dt + d\mathbf{b}(t) \\ \mathbf{x}(0, \rho) &= \mathbf{x}_0(\rho) \end{aligned} \quad (4.13)$$

where \mathbf{x} is the stochastic temperature, u is a heat source, \mathbf{b} is a noise process described at the end of this subsection, $B = 1$ is a constant input distributor⁵, and F is a time-invariant second-order partial differential operator. F is defined for every $\mathbf{x} \in \mathcal{D}(F)$ as

$$F\mathbf{x} \equiv \kappa \frac{\partial^2 \mathbf{x}}{\partial \rho^2} \text{ for } \kappa \in (0, \infty), \rho \in (0, 1) \quad (4.14)$$

for a domain described by

$$\mathcal{D}(F) = \left\{ \mathbf{x} \in \tilde{\mathbb{X}} : \mathbf{x}_\rho, \mathbf{x}_{\rho\rho} \in \tilde{\mathbb{X}}; \mathbf{x}_\rho(0) = 0 = \mathbf{x}_\rho(1) \right\} \quad (4.15)$$

and κ is the material thermal diffusivity constant. Since the stochastic temperature $\mathbf{x}(t, \rho)$, for $0 \leq \rho \leq 1$ is a continuous function over $[0, 1]$, it is a member of an infinite-dimensional vector space⁶. The stochastic state process, \mathbf{x} , is a collection of

⁵We have retained B to demonstrate how it flows through the development.

⁶An infinite-dimensional linear space requires an infinite number of basis vectors (in this case functions) to span the space. For example, the set of functions $y(\rho)$ defined on the interval $[0, 1]$ requires an infinite number of basis functions to span the space. Thus, the scalar function $y(\rho)$ has an infinite dimension.

random variables $\{\mathbf{x}(t) : t \in [t_0, t_f]\}$. The random variable $\mathbf{x}(t) \in \tilde{\mathbb{X}}$, is the same as defined in the previous section, thus, $\tilde{\mathbb{X}} = \mathbb{L}^2(\Omega, P; \mathbb{L}^2([0, 1], \mathbb{R}))$. The noise process⁷ \mathbf{b} is an \mathbb{X} -valued Wiener process⁸, described in Definition 61, with a time-invariant (constant) diffusion $Q \in \mathbb{R}$. The initial state \mathbf{x}_0 is an \mathbb{X} -valued Gaussian random process, see Definitions 59 and 51, with covariance operator $P_0 \in \mathcal{BLO}(\tilde{\mathbb{X}})$, where $\mathcal{BLO}(\tilde{\mathbb{X}})$ denotes the linear space of bounded linear operators (BLOs) on $\tilde{\mathbb{X}}$.

4.2.5 Equivalent Discrete-Time Dynamics Model. The equivalent discrete-time dynamics model for our scalar heat equation given in Equation (4.13) follows from Equation (3.224) given on page 3-73:

$$\mathbf{x}(t_{i+1}) = \Phi(t_{i+1} - t_i) \mathbf{x}(t_i) + B_d(t_i) u(t_i) + \mathbf{w}_d(t_i) \quad (4.16)$$

where $\Phi(t_{i+1} - t_i)$ is the state transition operator that we will define after we first discuss the equivalent discrete-time input distributor, B_d ,

$$B_d(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1} - s) B ds \quad (4.17)$$

provided $u(t)$ is a piece-wise constant function that is constant over each sample period, $\mathbf{G}_d(t_i)$ is not shown explicitly since it is an identity operator, and the equivalent discrete-time dynamics noise process \mathbf{w}_d is a zero-mean white Gaussian defined at each time t_i by

$$\mathbf{w}_d(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1} - s) d\mathbf{b}(s) \quad (4.18)$$

⁷In general, this noise process would be both a function of time and space; however, to make this example more tractable, it is assumed to be a time-varying, spatially-invariant function.

⁸The Wiener process is also known as Brownian motion (and thus the choice of notation \mathbf{b}); see Definition 61 on page 3-39. For an interesting discussion on the Brownian motion process, see for example, the book by Wiener, *et al.* [209]. In the derivative form of Equation (4.13), the $d\mathbf{b}/dt$ would represent the additive white noise – this is *not* the same as a Wiener process. Heuristically, we often treat the hypothetical time derivative of the Wiener process $d\mathbf{b}/dt$ as white Gaussian noise. In short, do not confuse this Brownian motion \mathbf{b} with a white Gaussian noise process!

with equivalent discrete-time dynamics noise covariance kernel operator, $\Sigma[\mathbf{w}_d(t_i), \mathbf{w}_d(t_k)] = 0$, for $t_i \neq t_k$, and covariance operator whenever $t_i = t_k$:

$$\Sigma[\mathbf{w}_d(t_i)] = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1} - s) Q \Phi^*(t_{i+1} - s) ds = Q_d(t_i) \quad (4.19)$$

The state transition operator Φ is, in general, a function of both time arguments, but here it is a function of a single parameter (the time difference) since F is time-invariant as can be seen in Equation (4.14), and is defined as⁹ [38]

$$[\Phi(t - t_0) \mathbf{x}_0](\rho) = \sum_{n=-\infty}^{\infty} e^{-\kappa n^2 \pi^2 (t-t_0)} \cos(n\pi\rho) \int_0^1 \mathbf{x}_0(\rho') \cos(n\pi\rho') d\rho' \quad (4.20)$$

For a discrete set of times, \mathbb{T} , the state transition operator Φ becomes

$$[\Phi(t_{i+1} - t_i) \mathbf{x}(t_i)](\rho) = \sum_{n=-\infty}^{\infty} e^{-\kappa n^2 \pi^2 (t_{i+1}-t_i)} \cos(n\pi\rho) \int_0^1 \mathbf{x}(t_i, \rho') \cos(n\pi\rho') d\rho' \quad (4.21)$$

4.2.6 Equivalent Infinite-Dimensional Discrete-Time Model. To summarize, we shall employ the results from Chapter III as a template for composing the dynamics and measurement model equations for this stochastic heat equation example. The results in Section 3.4 enable us to write down the important equations for the dynamics model that we have formally written in the introduction to this chapter. Since we want to use the ISKF to estimate the temperature optimally along the slender cylindrical rod, we require that our measurement model match the form of a generalized stochastic measurement model as given in Definition 76. We satisfied this requirement with a vector measurement model described in Section 4.2.3. Note that this vector measurement process, \mathbf{z} , is finite-dimensional — a special case of the general theory developed in Chapter III which will adequately serve to illustrate

⁹It is not a simple process to find the state transition (or evolution) operator for a time varying F ; see, for example, Engel and Nagel [48] for more evolution *operators* or Maybeck [129] for assistance in determining state transition *matrices*.

this research. Therefore, we may use the ISKF to estimate the scalar state process, \mathbf{x} , at times $\mathbb{T} = \{t_0, t_1, \dots, t_{\text{final}}\}$ using the dynamics equation

$$\mathbf{x}(t_{i+1}) = \Phi(t_{i+1} - t_i) \mathbf{x}(t_i) + B_d(t_i) u(t_i) + \mathbf{w}_d(t_i) \quad (4.16)$$

where $\Phi(t_{i+1} - t_i)$, $B_d(t_i)$, and $\mathbf{w}_d(t_i)$ are as previously described, for known control inputs $u(t_i)$, given noise-corrupted sampled-data vector measurements of the form

$$\mathbf{z}(t_i) = \mathbf{H} \mathbf{x}(t_i) + \mathbf{v}(t_i) \quad (4.22)$$

where the measurement transformation, \mathbf{H} , creates a column vector of M integrals over the state as previously given in Equation (4.11), and \mathbf{v} is the zero-mean white Gaussian noise process.

4.2.7 Essentially-Equivalent Finite-Dimensional Discrete-Time Model. In accordance with our stated methodology expressed by Figure 4.1 on page 4-5, we started with an infinite-dimensional continuous-time model, mapped it to the equivalent infinite-dimensional, discrete-time model, and then created an essentially-equivalent finite-dimensional, discrete-time model. We shall use the ISKF with a finite-dimensional approximation of the state function to estimate the state of the system optimally on a particular subspace described by an essentially-equivalent finite-dimensional, discrete-time model; in essence, we will be using the finite-dimensional sampled-data Kalman filter.

During the derivation of the ISKF we made no attempt to define the attributes of the various transformations, operators, and functions¹⁰ beyond what was needed to use the tools of functional analysis properly. However, in the example developed in this chapter, we have defined these transformations fully. Generally, it is not feasible

¹⁰Transformations, operators, and functions are all mappings, hence, we will often use the term transformations to refer to all of them when it is not important to distinguish between the various types of mappings.

to implement an infinite-dimensional transformation using a finite-dimensional computer algorithm with limited computational capabilities. Thus, our next task is to find appropriate matrix representations of the infinite-dimensional transformations on a finite-dimensional basis¹¹.

The following process is carried out for each of the transformations and operators present in the ISKF. Let T be a transformation acting on the state, $x \in \mathbb{X}$ and let $Tx \in \mathbb{Y}$. We shall approximate the state function by projecting the state onto a finite-dimensional subspace of \mathbb{X} , namely: $\mathcal{P}\mathbb{X} = \{\tilde{x} = \mathcal{P}x : \forall x \in \mathbb{X}\} \in \mathbb{R}^N$, where \mathcal{P} is a projection operator and \mathbb{R}^N is an N -dimensional Euclidean space. The approximate state is written as $\tilde{x} = \boldsymbol{\alpha}^T \boldsymbol{\beta}$, where $\boldsymbol{\alpha}$ is a vector of coefficients corresponding to the vector of basis elements, $\boldsymbol{\beta}$. For the approximate state, \tilde{x} , we have $T\tilde{x} \in \mathcal{P}\mathbb{Y}$. Evaluation of $T\tilde{x}$ yields an expression containing a finite-dimensional matrix representation of the transformation, denoted by $\tilde{\mathbf{T}}$. (Note that the tilde above \mathbf{T} is used to denote the finite-dimensional approximation of the infinite-dimensional function or transformation.) Hence, whenever the state function x is limited to a finite-dimensional subspace of the infinite-dimensional space, it reduces the dimensionality of the infinite-dimensional transformation such that we are left with a finite-dimensional matrix representation!

Admittedly, there are many ways to approximate the state function. Since the aim of this research is more in line with “demonstrating the concept” rather than “optimizing the methodology”, we have chosen to approximate the state function with the first N terms of a Fourier series expansion of the function — this will be discussed in more depth in the next section. Note that the finite-dimensional matrix representation of an infinite-dimensional transformation depends on the finite-dimensional approximation of the state function.

¹¹This process is conceptually analogous to reduced order filtering; in reduced order filtering, relatively unimportant states of the system “truth” model are removed or combined with other states in order to reduce the complexity of the model used to design a filter such that it becomes a more cost effective implementation [129, 175, 176, 71]

Thus, instead of propagating and updating the actual infinite-dimensional transformations and functions as defined in the previous sections, we will propagate and update matrices and vectors that represent these infinite-dimensional transformations and functions on a finite-dimensional subspace. In the end, our implemented algorithm will be nearly identical to the standard sampled-data Kalman filter.

For finite-dimensional approximations of the state, input, and noise functions, the system dynamics described by Equation (4.16) becomes

$$\tilde{\mathbf{x}}(t_{i+1}) = \Phi(t_{i+1} - t_i) \tilde{\mathbf{x}}(t_i) + B_d(t_i) \tilde{u}(t_i) + \tilde{\mathbf{w}}_d(t_i) \quad (4.23)$$

and Equation (4.22) for a given noise-corrupted sampled-data vector of measurements takes the form

$$\mathbf{z}(t_i) = \mathbf{H} \tilde{\mathbf{x}}(t_i) + \mathbf{v}(t_i) \quad (4.24)$$

Just as we often work calculus and algebra problems as far as possible before evaluating them for a particular numerical solution, we usually develop ancillary relationships prior to employing the characteristics of the finite-dimensional approximations. If we were to make the approximations first, then we would have to repeat many steps of the derivations each time the approximation changed. However, the results are the same, only the workload changes.

4.2.8 Approximating the Infinite-Dimensional State Via Projection. It is impractical to determine the exact (infinite-dimensional) temperature function x along the slender cylindrical rod using a finite number of sensors. Therefore, we desire a suitable approximation for x . Since we know that the function x is an element in a separable Hilbert space, it can be expanded in a Fourier series [154] without approximation for time t_i as

$$x(t_i, \rho) = \sum_{n=0}^{\infty} \langle x(t_i, \rho), \beta_n(\rho) \rangle \beta_n(\rho) = \sum_{n=0}^{\infty} \alpha_n(t_i) \beta_n(\rho) \quad (4.25)$$

with coefficients α_n defined as

$$\alpha_n(t_i) = \langle x(t_i, \rho), \beta_n(\rho) \rangle = \int_0^1 x(t_i, \rho) \beta_n(\rho) d\rho \quad (4.26)$$

and the orthonormal set $\mathbb{B}_\infty = \{\beta_0(\rho), \beta_1(\rho), \dots\}$, defined as

$$\beta_n(\rho) \triangleq \begin{cases} 1, & n = 0 \\ \sqrt{2} \cos(n\pi\rho), & n > 0 \end{cases} \quad (4.27)$$

forms an orthonormal basis for this Hilbert space. Thus, in this formulation of x , we may have to compute a countably infinite number of coefficients at each time t_i in order to estimate the temperature along the rod accurately. Each coefficient, α_n , is computed by taking the inner product of the state with the corresponding member of the basis, say β_n , then

$$\langle x(t_i, \rho), \beta_n(\rho) \rangle = \left\langle \sum_{m=0}^{\infty} \alpha_m(t_i) \beta_m(\rho), \beta_n(\rho) \right\rangle = \sum_{m=0}^{\infty} \alpha_m(t_i) \langle \beta_m(\rho), \beta_n(\rho) \rangle \quad (4.28)$$

The orthogonality property of an orthogonal basis reduces the inner product to

$$\langle \beta_m, \beta_n \rangle = \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases} = \delta_{mn} \quad (4.29)$$

where δ_{mn} is known as the Kronecker delta. Hence we can write Equation (4.28) as

$$\langle x(t_i, \rho), \beta_n(\rho) \rangle = \sum_{m=0}^{\infty} \alpha_m(t_i) \delta_{mn}(\rho) \quad (4.30)$$

The right-hand side is clearly $\alpha_n(t_i)$ since δ_{mn} is zero for every term except $m = n$. Therefore, Equation (4.26) is true and Equation (4.25) has been completely justified.

The infinite-dimensional scalar state function x can be well approximated by projecting the infinite-dimensional function, x , onto a finite-dimensional subspace by truncating the Fourier series expansion after N terms. This makes good sense from an engineering point of view — the information in the high frequency terms is often dominated by noise; hence we are effectively low-pass filtering the state with an ideal lowpass filter. Thus, projecting x onto a finite N -dimensional space is accomplished by

$$\tilde{x}(t_i, \rho) = [\mathcal{P}x](t_i, \rho) = \mathcal{P} \left[\sum_{n=0}^{\infty} \alpha_n(t_i) \beta_n(\rho) \right] \quad (4.31)$$

where \mathcal{P} is the projection operator, the coefficients $\alpha_0, \alpha_1, \dots$ are as defined as above in Equation (4.26), and a basis for this subspace is given in Equation (4.27). Thus,

$$\tilde{x}(t_i, \rho) = \sum_{n=0}^{N-1} \alpha_n(t_i) \beta_n(\rho) \quad (4.32)$$

and the basis is denoted by

$$\mathbb{B}_N = \{\beta_0(\rho), \beta_1(\rho), \dots, \beta_{N-1}(\rho)\} \quad (4.33)$$

A more convenient form for Equation (4.32), using vector multiplication, is

$$\tilde{x}(t_i, \rho) = \boldsymbol{\alpha}^T(t_i) \boldsymbol{\beta}(\rho) \quad (4.34)$$

where $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}(\rho)$ are defined as column vectors:

$$\boldsymbol{\alpha}(t_i) \equiv \begin{bmatrix} \alpha_0(t_i) & \alpha_1(t_i) & \cdots & \alpha_{N-1}(t_i) \end{bmatrix}^T \quad (4.35)$$

$$\boldsymbol{\beta}(\rho) \equiv \begin{bmatrix} \beta_0(\rho) & \beta_1(\rho) & \cdots & \beta_{N-1}(\rho) \end{bmatrix}^T \quad (4.36)$$

Additionally, the coordinate vector of \tilde{x} with respect to the basis \mathbb{B}_N is [85]

$$\tilde{x}(t_i, \rho) = [\boldsymbol{\alpha}(t_i)]_{\mathbb{B}_N} \quad (4.37)$$

Furthermore, the state estimation error is defined by

$$\mathbf{e}(t_i^-) \triangleq \mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-) \quad (4.38)$$

and the finite-dimensional approximation of the state estimation error is given by a truncated Fourier series expansion as:

$$\tilde{\mathbf{e}}(t_i^-, \rho) = \mathcal{P}\mathbf{e}(t_i^-, \rho) = \sum_{n=0}^{N-1} \epsilon_n(t_i^-) \beta_n(\rho) \quad (4.39)$$

where the coefficients are determined by

$$\epsilon_n(t_i^-) = \langle \mathbf{e}(t_i^-), \beta_n \rangle \quad (4.40)$$

Remark: In most applications, researchers have discretized the state x in the spatial domain. In contrast, we have discretized the state in the spatial-frequency domain and truncated the higher frequencies that are often dominated by noise. If truncation performed by this projection operator causes too much ringing or other undesirable effects in the spatial domain representation of the state, then perhaps additional filtering techniques (or an entirely different technique) may be needed in order to achieve good performance. Finally, by choosing another subset of basis elements or by using a different basis altogether, we will, in general, produce different matrix representations. Hence, the basis and projection operator employed can affect the speed, cost, and effectiveness of the calculation as well as the accuracy of the estimation process.

4.3 Kalman Filtering Algorithm

An inspection of the sampled-data Kalman filtering algorithm discussed in Section 2.3.2 reveals the following list of transformations

$$\Phi(t_i, t_{i-1}), \mathbf{B}_d(t_i), \mathbf{G}_d(t_i), \mathbf{Q}_d(t_i), \mathbf{P}(t_i^-), \mathbf{H}(t_i), \mathbf{R}(t_i), \mathbf{P}(t_i^+), \mathbf{A}(t_i), \mathbf{K}(t_i)$$

used by the algorithm to perform its optimal estimation of the state; we will often refer to these transformations as the *components* of a Kalman filter¹². Before we derive the matrix representations for the transformations using a finite-dimensional approximation of the state function, we shall first find the state transition operator adjoint and the measurement distributor transformation adjoint, and perform some preliminary work regarding the residual covariance matrix and the Kalman filter gain transformation.

4.3.1 The State Transition Operator Adjoint Φ^ .* The state transition operator maps both a random state function and its realizations. For ease of development, we shall begin by determining the state transition adjoint operator Φ^* applied to a realization of the stochastic (state) temperature function. The following fundamental equation relates how $\Phi \in \mathcal{BLO}(\mathbb{X})$ is related to its adjoint¹³ via the inner product, $\langle \cdot, \cdot \rangle$, at time t_i [154]

$$\langle \Phi(\Delta t_i) x(t_{i-1}), y(t_{i-1}) \rangle = \langle x(t_{i-1}), \Phi^*(\Delta t_i) y(t_{i-1}) \rangle \quad (4.41)$$

for every $x, y \in \mathbb{X} = \mathbb{L}_{[0,1]}^2$ taken over the Hilbert space of absolutely square integrable real-valued functions. By definition,

$$\langle \Phi(\Delta t_i) x(t_{i-1}), y(t_{i-1}) \rangle = \int_0^1 [\Phi(\Delta t_i) x(t_{i-1})](\rho) y(t_{i-1}, \rho) d\rho \quad (4.42)$$

¹²In this example, $\mathbf{G}_d(t_i)$ is an identity operator. Without loss of generality, we could have let $\mathbf{G}_d(t_i)$ be an identity operator in previous discussions.

¹³Let \mathbb{X} and \mathbb{Y} be Hilbert spaces, then for every pair of elements $x \in \mathbb{X}$ and $y \in \mathbb{Y}$, the linear transformation $\Psi \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$ is related to its adjoint via the inner product $\langle \Psi x, y \rangle_{\mathbb{Y}} = \langle x, \Psi^* y \rangle_{\mathbb{X}}$ [154].

Substituting in for $[\Phi(\Delta t_i) x(t_{i-1})](\rho)$ using Equation (4.21) yields

$$\begin{aligned} & \langle \Phi(\Delta t_i) x(t_{i-1}), y(t_{i-1}) \rangle \\ &= \int_0^1 \left[\sum_{n=-\infty}^{\infty} e^{-\kappa n^2 \pi^2 \Delta t_i} \cos(n\pi\rho) \int_0^1 x(t_{i-1}, \rho') \cos(n\pi\rho') d\rho' \right] y(t_{i-1}, \rho) d\rho \end{aligned} \quad (4.43)$$

$$= \int_0^1 \sum_{n=-\infty}^{\infty} e^{-\kappa n^2 \pi^2 \Delta t_i} y(t_{i-1}, \rho) \cos(n\pi\rho) \int_0^1 x(t_{i-1}, \rho') \cos(n\pi\rho') d\rho' d\rho \quad (4.44)$$

This inner product is finite because it is defined on the space of absolutely square integrable functions, thus, the infinite sum in the integrand converges. Therefore, by the Weierstrauss M-Test [7] we can interchange the ordering and pull the infinite sum outside of the integral to get

$$\begin{aligned} & \langle \Phi(\Delta t_i) x(t_{i-1}), y(t_{i-1}) \rangle \\ &= \sum_{n=-\infty}^{\infty} e^{-\kappa n^2 \pi^2 \Delta t_i} \int_0^1 y(t_{i-1}, \rho) \cos(n\pi\rho) d\rho \int_0^1 x(t_{i-1}, \rho') \cos(n\pi\rho') d\rho' \end{aligned} \quad (4.45)$$

Next, we interchange the order of integrations by invoking the Fubini-Tonelli theorem [194, 141, 7, 66] since $y(t_{i-1}, \rho) \cos(n\pi\rho) x(t_{i-1}, \rho') \cos(n\pi\rho')$ is absolutely integrable. Thus, Equation (4.45) becomes

$$\begin{aligned} & \langle \Phi(\Delta t_i) x(t_{i-1}), y(t_{i-1}) \rangle \\ &= \int_0^1 x(t_{i-1}, \rho') \left[\sum_{n=-\infty}^{\infty} e^{-\kappa n^2 \pi^2 \Delta t_i} \cos(n\pi\rho') \int_0^1 y(t_{i-1}, \rho) \cos(n\pi\rho) d\rho \right] d\rho' \end{aligned} \quad (4.46)$$

Per Equation (4.21), the term inside the large square brackets in Equation (4.46) is

$$\sum_{n=-\infty}^{\infty} e^{-\kappa n^2 \pi^2 \Delta t_i} \cos(n\pi\rho') \int_0^1 y(t_{i-1}, \rho) \cos(n\pi\rho) d\rho = [\Phi(\Delta t_i) y(t_{i-1})](\rho') \quad (4.47)$$

Substituting this into Equation (4.46) gives

$$\langle \Phi(\Delta t_i) x(t_{i-1}), y(t_{i-1}) \rangle = \int_0^1 x(t_{i-1}, \rho') [\Phi(\Delta t_i) y(t_{i-1})](\rho') d\rho' \quad (4.48)$$

A comparison of the right-hand sides of Equations (4.41) and (4.48) yields: $\Phi^* = \Phi$. When an operator is equal to its adjoint, the operator is termed *self-adjoint*. For a matrix representation of the operator, the conjugate transpose of the matrix yields the adjoint. Hence the matrix representation of a self-adjoint operator is symmetric if the matrix is real, and Hermitian symmetric if complex.

4.3.2 The Measurement Distributor Transformation Adjoint \mathbf{H}^ .* In order to find the filter-computed error covariance, $\mathbf{A} = \mathbf{H}\mathbf{P}\mathbf{H}^* + \mathbf{R}$, we must first determine \mathbf{H}^* , the adjoint of \mathbf{H} . For a given $\omega \in \Omega$, $\mathbf{x}(\omega) = x \in \mathbb{X} = \mathbb{L}_{[0,1]}^2$. The transformation \mathbf{H} is defined for arbitrary x by

$$[\mathbf{H}x](t_i) \triangleq \begin{bmatrix} \int_0^{1/M} x(t_i, \rho) d\rho \\ \vdots \\ \int_{(M-1)/M}^1 x(t_i, \rho) d\rho \end{bmatrix} \in \mathbb{R}^M \quad (4.49)$$

for some time $t_i \in \{t_1, \dots, t_{\text{final}}\}$. Thus \mathbf{H} is a vector of linear functionals. For any $t_i \in \{t_1, \dots, t_{\text{final}}\}$, $x \in \mathbb{L}_{[0,1]}^2$, $\mathbf{y} \in \mathbb{R}^M$, we can use the following definition of the adjoint, an equality of inner products

$$\langle \mathbf{H}x, \mathbf{y} \rangle_{\mathbb{R}^M} = \langle x, \mathbf{H}^* \mathbf{y} \rangle_{\mathbb{L}^2} \quad (4.50)$$

to determine the adjoint \mathbf{H}^* . The left-hand side of Equation (4.50) is given by

$$\langle \mathbf{H}x, \mathbf{y} \rangle_{\mathbb{R}^M} = \begin{bmatrix} \int_0^{1/M} x(t_i, \rho) d\rho & \cdots & \int_{(M-1)/M}^1 x(t_i, \rho) d\rho \end{bmatrix} \mathbf{y}(t_i)$$

Performing the inner product yields

$$\begin{aligned} & \langle \mathbf{H}x, \mathbf{y} \rangle_{\mathbb{R}^M} \\ &= \left[\int_0^{1/M} x(t_i, \rho) d\rho \right] y_1(t_i) + \cdots + \left[\int_{(M-1)/M}^1 x(t_i, \rho) d\rho \right] y_M(t_i) \end{aligned} \quad (4.51)$$

$$= \int_0^{1/M} x(t_i, \rho) y_1(t_i) d\rho + \cdots + \int_{(M-1)/M}^1 x(t_i, \rho) y_M(t_i) d\rho \quad (4.52)$$

Furthermore, the right-hand side of Equation (4.50) is

$$\langle x, \mathbf{H}^* \mathbf{y} \rangle_{\mathbb{L}^2} = \int_0^1 x(t_i, \rho) [\mathbf{H}^* \mathbf{y}](t_i, \rho) d\rho \quad (4.53)$$

$$= \sum_{m=1}^M \int_{(m-1)/M}^{m/M} x(t_i, \rho) [\mathbf{H}^* \mathbf{y}](t_i, \rho) d\rho \quad (4.54)$$

So,

$$\sum_{m=1}^M \int_{(m-1)/M}^{m/M} x(t_i, \rho) y_m(t_i) d\rho = \sum_{m=1}^M \int_{(m-1)/M}^{m/M} x(t_i, \rho) [\mathbf{H}^* \mathbf{y}](t_i, \rho) d\rho \quad (4.55)$$

and for Equation (4.55) to be true, it follows that the summands must be equal for each m so

$$\int_{(m-1)/M}^{m/M} x(t_i, \rho) y_m(t_i) d\rho = \int_{(m-1)/M}^{m/M} x(t_i, \rho) [\mathbf{H}^* \mathbf{y}](t_i, \rho) d\rho \quad (4.56)$$

Thus, the integrands over each subinterval are equal almost everywhere in ρ

$$x(t_i, \rho) y_m(t_i) = x(t_i, \rho) [\mathbf{H}^* \mathbf{y}](t_i, \rho) \quad (4.57)$$

Therefore $[\mathbf{H}^* \mathbf{y}](t_i, \rho) = y_m(t_i)$ almost everywhere in ρ for every $m = \{1, \dots, M\}$, and any fixed t_i . So, for fixed t_i , the *measurement distributor transformation adjoint* \mathbf{H}^* transforms a constant vector $\mathbf{y}(t_i) \in \mathbb{R}^M$ into a piece-wise constant $\mathbb{L}_{[0,1]}^2$ function

$$[\mathbf{H}^* \mathbf{y}](t_i, \rho) = y_m(t_i), \text{ for } \frac{m-1}{M} \leq \rho \leq \frac{m}{M} \quad (4.58)$$

4.3.3 The Residual Covariance Matrix $\mathbf{A}(t_i)$. The residual covariance operator $\mathbf{A}(t_i) = \mathbf{H} P(t_i^-) \mathbf{H}^* + \mathbf{R}(t_i)$, was previously defined in Equation (3.265). We will show that $\mathbf{H} P(t_i^-) \mathbf{H}^*$ produces an $M \times M$ real-valued matrix; hence \mathbf{A} is an $M \times M$ real-valued matrix. For an arbitrary sample of \mathbf{z} at time t_i , the measurement vector $\mathbf{z}_i \in \mathbb{R}^M$

$$\mathbf{A}(t_i) \mathbf{z}_i = [\mathbf{H} P(t_i^-) \mathbf{H}^* + \mathbf{R}(t_i)] \mathbf{z}_i \quad (4.59)$$

$$= \mathbf{H} P(t_i^-) \mathbf{H}^* \mathbf{z}_i + \mathbf{R}(t_i) \mathbf{z}_i \quad (4.60)$$

The second term is just matrix multiplication and requires no further discussion at this time. On the other hand, the first term involves two transformations and an operator and will require an extensive development. We shall begin by explicitly calling out the form of error covariance operator $P(t_i^-)$ applied to the function $\mathbf{H}^* \mathbf{z}_i \in \mathbb{L}_{[0,1]}^2$ by

$$P(t_i^-) \mathbf{H}^* \mathbf{z}_i = E\{[\mathbf{e}(t_i^-) \diamond \mathbf{e}(t_i^-)] | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\} \mathbf{H}^* \mathbf{z}_i \quad (4.61)$$

where $E\{\cdot\}$ is the conditional expectation operator¹⁴ and the outer product operator, \diamond , is defined in Definition 14 in Chapter III. Next, we move $\mathbf{H}^* \mathbf{z}_i$ into the expectation

$$P(t_i^-) \mathbf{H}^* \mathbf{z}_i = E\{[\mathbf{e}(t_i^-) \diamond \mathbf{e}(t_i^-)] \mathbf{H}^* \mathbf{z}_i\} \quad (4.62)$$

and then use the definition of the outer product given in Equation (3.6) to obtain

$$P(t_i^-) \mathbf{H}^* \mathbf{z}_i = E\{\mathbf{e}(t_i^-) \langle \mathbf{e}(t_i^-), \mathbf{H}^* \mathbf{z}_i \rangle_{\mathbb{L}^2}\} \quad (4.63)$$

where the inner product is defined on the \mathbb{L}^2 space of functions for interval $[0, 1]$. The inner product on the right-hand side can be equivalently expressed using the relationship employed to define the adjoint of the measurement distributor transformation. We get

$$\langle \mathbf{e}(t_i^-), \mathbf{H}^* \mathbf{z}_i \rangle_{\mathbb{L}^2} = \langle \mathbf{H} \mathbf{e}(t_i^-), \mathbf{z}_i \rangle_{\mathbb{R}^M} \quad (4.64)$$

and since the inner product of two vectors in \mathbb{R}^M is

$$\langle \mathbf{H} \mathbf{e}(t_i^-), \mathbf{z}_i \rangle_{\mathbb{R}^M} = [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{z}_i \quad (4.65)$$

¹⁴For ease of notation, we will suppress the explicit conditioning during the majority of the development.

thus

$$P(t_i^-) \mathbf{H}^* \mathbf{z}_i = E\{\mathbf{e}(t_i^-) [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{z}_i\} \quad (4.66)$$

Next we apply \mathbf{H} to $P(t_i^-) \mathbf{H}^* \mathbf{z}_i$ to yield

$$\mathbf{H} P(t_i^-) \mathbf{H}^* \mathbf{z}_i = \begin{bmatrix} \int_0^{1/M} E\{\mathbf{e}(t_i^-, \rho) [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{z}_i\} d\rho \\ \vdots \\ \int_{(M-1)/M}^1 E\{\mathbf{e}(t_i^-, \rho) [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{z}_i\} d\rho \end{bmatrix} \quad (4.67)$$

where we now include the linear spatial variable ρ dependence when applicable.

Next, we pull the expectation operator out of the integrals

$$\mathbf{H} P(t_i^-) \mathbf{H}^* \mathbf{z}_i = \begin{bmatrix} E \left\{ \int_0^{1/M} \mathbf{e}(t_i^-, \rho) [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{z}_i d\rho \right\} \\ \vdots \\ E \left\{ \int_{(M-1)/M}^1 \mathbf{e}(t_i^-, \rho) [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{z}_i d\rho \right\} \end{bmatrix} \quad (4.68)$$

and then out of the array

$$\mathbf{H} P(t_i^-) \mathbf{H}^* \mathbf{z}_i = E \begin{bmatrix} \int_0^{1/M} \mathbf{e}(t_i^-, \rho) [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{z}_i d\rho \\ \vdots \\ \int_{(M-1)/M}^1 \mathbf{e}(t_i^-, \rho) [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{z}_i d\rho \end{bmatrix} \quad (4.69)$$

since a vector of expectations is equivalent to the expectation of the vector. For the m th element, expanding $\mathbf{H} \mathbf{e}(t_i^-)$ yields

$$\begin{aligned} & \int_{(m-1)/M}^{m/M} \mathbf{e}(t_i^-, \rho) [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{z}_i d\rho \\ &= \int_{(m-1)/M}^{m/M} \mathbf{e}(t_i^-, \rho) \left[\int_0^{1/M} \mathbf{e}(t_i^-, \rho') d\rho' \cdots \int_{(M-1)/M}^1 \mathbf{e}(t_i^-, \rho') d\rho' \right] \mathbf{z}_i d\rho \quad (4.70) \end{aligned}$$

$$= \int_{(m-1)/M}^{m/M} \mathbf{e}(t_i^-, \rho) d\rho \left[\int_0^{1/M} \mathbf{e}(t_i^-, \rho') d\rho' \cdots \int_{(M-1)/M}^1 \mathbf{e}(t_i^-, \rho') d\rho' \right] \mathbf{z}_i \quad (4.71)$$

where the second equality follows from the fact that the integrations are separable. For notational convenience, we define the *integrated error* as

$$\eta_m(t_i^-) \triangleq \int_{(m-1)/M}^{m/M} \mathbf{e}(t_i^-, \rho) d\rho \quad (4.72)$$

so that Equation (4.71) for the m th element becomes

$$\begin{aligned} & \int_{(m-1)/M}^{m/M} \mathbf{e}(t_i^-, \rho) [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{z}_i d\rho \\ &= \eta_m(t_i^-) \begin{bmatrix} \eta_1(t_i^-) & \cdots & \eta_M(t_i^-) \end{bmatrix} \mathbf{z}_i \end{aligned} \quad (4.73)$$

$$= \begin{bmatrix} \eta_m(t_i^-) \eta_1(t_i^-) & \cdots & \eta_m(t_i^-) \eta_M(t_i^-) \end{bmatrix} \mathbf{z}_i \quad (4.74)$$

Using Equation (4.74), $\mathbf{H} P(t_i^-) \mathbf{H}^* \mathbf{z}_i$ becomes

$$\mathbf{H} P(t_i^-) \mathbf{H}^* \mathbf{z}_i = E \begin{bmatrix} \begin{bmatrix} \eta_1(t_i^-) \eta_1(t_i^-) & \cdots & \eta_1(t_i^-) \eta_M(t_i^-) \end{bmatrix} \mathbf{z}_i \\ \vdots \\ \begin{bmatrix} \eta_M(t_i^-) \eta_1(t_i^-) & \cdots & \eta_M(t_i^-) \eta_M(t_i^-) \end{bmatrix} \mathbf{z}_i \end{bmatrix} \quad (4.75)$$

$$= E \begin{bmatrix} \eta_1(t_i^-) \eta_1(t_i^-) & \cdots & \eta_1(t_i^-) \eta_M(t_i^-) \\ \vdots & \ddots & \vdots \\ \eta_M(t_i^-) \eta_1(t_i^-) & \cdots & \eta_M(t_i^-) \eta_M(t_i^-) \end{bmatrix} \mathbf{z}_i \quad (4.76)$$

Applying the expectation to the individual elements of the matrix yields

$$\mathbf{H} P(t_i^-) \mathbf{H}^* \mathbf{z}_i = \begin{bmatrix} E\{\eta_1(t_i^-) \eta_1(t_i^-)\} & \cdots & E\{\eta_1(t_i^-) \eta_M(t_i^-)\} \\ \vdots & \ddots & \vdots \\ E\{\eta_M(t_i^-) \eta_1(t_i^-)\} & \cdots & E\{\eta_M(t_i^-) \eta_M(t_i^-)\} \end{bmatrix} \mathbf{z}_i \quad (4.77)$$

where the m th element is

$$E\{\eta_m(t_i^-) \eta_n(t_i^-)\} = E \left\{ \int_{(m-1)/M}^{m/M} \mathbf{e}(t_i^-, \rho) d\rho \int_{(n-1)/M}^{n/M} \mathbf{e}(t_i^-, \rho') d\rho' \right\} \quad (4.78)$$

$$= E \left\{ \int_{(m-1)/M}^{m/M} \int_{(n-1)/M}^{n/M} \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') d\rho' d\rho \right\} \quad (4.79)$$

$$= \int_{(m-1)/M}^{m/M} \int_{(n-1)/M}^{n/M} E \{ \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') \} d\rho' d\rho \quad (4.80)$$

By the operator identity for arbitrary \mathbf{z}_i ,

$$\mathbf{H} P(t_i^-) \mathbf{H}^* = \begin{bmatrix} E\{\eta_1(t_i^-) \eta_1(t_i^-)\} & \cdots & E\{\eta_1(t_i^-) \eta_M(t_i^-)\} \\ \vdots & \ddots & \vdots \\ E\{\eta_M(t_i^-) \eta_1(t_i^-)\} & \cdots & E\{\eta_M(t_i^-) \eta_M(t_i^-)\} \end{bmatrix} \quad (4.81)$$

and therefore the filter-computed error covariance operator is represented by a real $M \times M$ matrix

$$\mathbf{A}(t_i) = \begin{bmatrix} E\{\eta_1(t_i^-) \eta_1(t_i^-)\} & \cdots & E\{\eta_1(t_i^-) \eta_M(t_i^-)\} \\ \vdots & \ddots & \vdots \\ E\{\eta_M(t_i^-) \eta_1(t_i^-)\} & \cdots & E\{\eta_M(t_i^-) \eta_M(t_i^-)\} \end{bmatrix} + \mathbf{R}(t_i) \quad (4.82)$$

In this subsection we began with an M -vector of numbers, \mathbf{z}_i , and then transformed it into a function using \mathbf{H}^* . Application of the error covariance operator, $P(t_i^-)$, to $\mathbf{H}^* \mathbf{z}_i$ returned a modified function. Finally, we transformed the function $P(t_i^-) \mathbf{H}^* \mathbf{z}_i$ with \mathbf{H} to create another M -vector of numbers. The composite operator: $\mathbf{H} P(t_i^-) \mathbf{H}^*$ is thus a matrix of real numbers.

4.3.4 *The Kalman Gain Transformation K .* The Kalman gain transformation $K \in \mathcal{LT}(\mathbb{Z}, \mathbb{X})$, where $\mathbb{Z} = \mathbb{R}^M$ and $\mathbb{X} = \mathbb{L}_{[0,1]}^2$ for some time t_i ,

$$K(t_i) = P(t_i^-) \mathbf{H}^* \mathbf{A}^{-1}(t_i) \quad (4.83)$$

operates on the vector residual \mathbf{r} at time t_i

$$\mathbf{r}(t_i) = \mathbf{z}_i - \mathbf{H} \hat{\mathbf{x}}(t_i^-) \quad (4.84)$$

which for this problem is a vector of real scalars, i.e., $\mathbf{r}(t_i) \in \mathbb{R}^M$. Apply the gain to the residual

$$[K\mathbf{r}](t_i) = P(t_i^-) \mathbf{H}^* \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i) \quad (4.85)$$

$$= E\{\mathbf{e}(t_i^-) \diamond \mathbf{e}(t_i^-)\} \mathbf{H}^* \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i) \quad (4.86)$$

$$= E\{[\mathbf{e}(t_i^-) \diamond \mathbf{e}(t_i^-)] \mathbf{H}^* \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i)\} \quad (4.87)$$

where the expectation $E\{\mathbf{e}(t_i^-) \diamond \mathbf{e}(t_i^-)\}$ is actually a conditional expectation: $E\{[\mathbf{e}(t_i^-) \diamond \mathbf{e}(t_i^-)] | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\}$. Per the development in Section 4.3.2, which began on page 4-18, note that $\mathbf{H}^* \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i)$ is an $\mathbb{L}_{[0,1]}^2$ function. Using the definition of the function outer product in Equation (3.6), we get

$$[K\mathbf{r}](t_i) = E\{\mathbf{e}(t_i^-) \langle \mathbf{e}(t_i^-), \mathbf{H}^* \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i) \rangle_{\mathbb{L}^2}\} \quad (4.88)$$

$$= E\{\mathbf{e}(t_i^-) \langle \mathbf{H} \mathbf{e}(t_i^-), \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i) \rangle_{\mathbb{R}^M}\} \quad (4.89)$$

$$= E\{\mathbf{e}(t_i^-) [\mathbf{H} \mathbf{e}(t_i^-)]^T \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i)\} \quad (4.90)$$

where the second line follows from the equivalence of inner products in the definition of the adjoint and the third line is by definition of the inner product for vectors.

Expanding $[\mathbf{H} \mathbf{e}(t_i^-)]^T$ into a vector yields

$$[K\mathbf{r}](t_i) = E \left\{ \mathbf{e}(t_i^-, \rho) \begin{bmatrix} \int_0^{1/M} \mathbf{e}(t_i^-, \rho') d\rho' \\ \vdots \\ \int_{(M-1)/M}^1 \mathbf{e}(t_i^-, \rho') d\rho' \end{bmatrix}^T \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i) \right\} \quad (4.91)$$

$$= E \left\{ \begin{bmatrix} \int_0^{1/M} \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') d\rho' \\ \vdots \\ \int_{(M-1)/M}^1 \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') d\rho' \end{bmatrix}^T \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i) \right\} \quad (4.92)$$

and then moving the expectation operator into the row vector

$$[K\mathbf{r}](t_i) = \begin{bmatrix} E \left\{ \int_0^{1/M} \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') d\rho' \right\} \\ \vdots \\ E \left\{ \int_{(M-1)/M}^1 \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') d\rho' \right\} \end{bmatrix}^T \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i) \quad (4.93)$$

$$= \begin{bmatrix} \int_0^{1/M} E \left\{ \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') \right\} d\rho' \\ \vdots \\ \int_{(M-1)/M}^1 E \left\{ \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') \right\} d\rho' \end{bmatrix}^T \mathbf{A}^{-1}(t_i) \mathbf{r}(t_i) \quad (4.94)$$

where the expectation and integration operations commute and we note that the integrand is related to the error covariance. Since Equation (4.94) applies for all residuals $\mathbf{r}(t_i)$, by operator identity, the Kalman gain transformation $K(t_i)$ is a row vector of real numbers defined as

$$K(t_i) = \begin{bmatrix} \int_0^{1/M} E \left\{ \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') \right\} d\rho' \\ \vdots \\ \int_{(M-1)/M}^1 E \left\{ \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') \right\} d\rho' \end{bmatrix}^T \mathbf{A}^{-1}(t_i) \quad (4.95)$$

While the gain is a finite-dimensional vector of numbers, we still must address the integrals since they are, in general, infinite-dimensional operators. In Section 4.3.11, we will show that, for the basis that we chose in Section 4.2.8, we can analytically evaluate the integrals and store a vector of numbers to act as the approximate measurement distributor matrix.

4.3.5 The State Transition Matrix $\tilde{\Phi}(\Delta t_{i+1})$. To find the matrix representation of our state transition operator, denoted as $\tilde{\Phi}(\Delta t_{i+1})$ because it requires a finite-dimensional approximation of that operator (and thus the tilde over Φ versus no tilde over Φ), we evaluate $\Phi(\Delta t_{i+1}) \tilde{x}(t_i)$, where $\Delta t_{i+1} = t_{i+1} - t_i$. Using the defining relationship for the state transition operator, Equation (4.21), we obtain

$$[\Phi(\Delta t_{i+1}) \tilde{x}(t_i)](\rho) = \sum_{n=-\infty}^{\infty} e^{-\kappa n^2 \pi^2 \Delta t_{i+1}} \cos(n\pi\rho) \int_0^1 \cos(n\pi\rho') \tilde{x}(t_i, \rho') d\rho' \quad (4.96)$$

The evenness property of the summand allows us to double the sum over the positive indices and separately compute the $n = 0$ term, thus:

$$\begin{aligned} & [\Phi(\Delta t_{i+1}) \tilde{x}(t_i)](\rho) \\ &= \sum_{n=1}^{\infty} 2e^{-\kappa n^2 \pi^2 \Delta t_{i+1}} \cos(n\pi\rho) \int_0^1 \cos(n\pi\rho') \tilde{x}(t_i, \rho') d\rho' + \int_0^1 \tilde{x}(t_i, \rho') d\rho' \end{aligned} \quad (4.97)$$

Simplifying the first term yields

$$\begin{aligned} & \sum_{n=1}^{\infty} 2e^{-\kappa n^2 \pi^2 \Delta t_{i+1}} \cos(n\pi\rho) \int_0^1 \cos(n\pi\rho') \tilde{x}(t_i, \rho') d\rho' \\ &= \sum_{n=1}^{\infty} 2e^{-\kappa n^2 \pi^2 \Delta t_{i+1}} \cos(n\pi\rho) \int_0^1 \cos(n\pi\rho') \sum_{m=0}^{N-1} \alpha_m(t_i) \beta_m(\rho') d\rho' \end{aligned} \quad (4.98)$$

$$= \sum_{n=1}^{\infty} 2e^{-\kappa n^2 \pi^2 \Delta t_{i+1}} \frac{\beta_n(\rho)}{\sqrt{2}} \int_0^1 \frac{\beta_n(\rho')}{\sqrt{2}} \sum_{m=0}^{N-1} \alpha_m(t_i) \beta_m(\rho') d\rho' \quad (4.99)$$

$$= \sum_{n=1}^{\infty} e^{-\kappa n^2 \pi^2 \Delta t_{i+1}} \beta_n(\rho) \sum_{m=0}^{N-1} \alpha_m(t_i) \int_0^1 \beta_n(\rho') \beta_m(\rho') d\rho' \quad (4.100)$$

Recall that

$$\int_0^1 \beta_n(\rho') \beta_m(\rho') d\rho' = \langle \beta_n, \beta_m \rangle = \delta_{mn} \quad (4.101)$$

Thus the first term of Equation (4.97) becomes

$$\begin{aligned} & \sum_{n=1}^{\infty} 2e^{-\kappa n^2 \pi^2 \Delta t_{i+1}} \cos(n\pi\rho) \int_0^1 \cos(n\pi\rho') \tilde{x}(t_i, \rho') d\rho' \\ &= \sum_{n=1}^{\infty} e^{-\kappa n^2 \pi^2 \Delta t_{i+1}} \beta_n(\rho) \sum_{m=0}^{N-1} \alpha_m(t_i) \delta_{mn} \end{aligned} \quad (4.102)$$

$$= \sum_{n=1}^{\infty} \sum_{m=0}^{N-1} e^{-\kappa n^2 \pi^2 \Delta t_{i+1}} \beta_n(\rho) \alpha_m(t_i) \delta_{mn} \quad (4.103)$$

$$= \sum_{m=1}^{N-1} e^{-\kappa m^2 \pi^2 \Delta t_{i+1}} \alpha_m(t_i) \beta_m(\rho) \quad (4.104)$$

where we evaluated the Kronecker delta to obtain the last line. The second term in Equation (4.97) is evaluated as

$$\int_0^1 \tilde{x}(t_i, \rho') d\rho' = \int_0^1 \sum_{m=0}^{N-1} \alpha_m(t_i) \beta_m(\rho') d\rho' \quad (4.105)$$

$$= \sum_{m=0}^{N-1} \alpha_m(t_i) \int_0^1 \beta_m(\rho') d\rho' \quad (4.106)$$

$$= \alpha_0(t_i) \quad (4.107)$$

since the integral is zero for every index $m > 0$ and one when $m = 0$. Therefore, Equation (4.97) becomes

$$[\Phi(\Delta t_{i+1}) \tilde{x}(t_i)](\rho) = \alpha_0(t_i) + \sum_{m=1}^{N-1} e^{-\kappa m^2 \pi^2 \Delta t_{i+1}} \alpha_m(t_i) \beta_m(\rho) \quad (4.108)$$

$$= \sum_{m=0}^{N-1} e^{-\kappa m^2 \pi^2 \Delta t_{i+1}} \alpha_m(t_i) \beta_m(\rho) \quad (4.109)$$

where the base $\alpha_0(t_i)$ term rejoined the summation since both $\beta_0(\rho)$ and e^0 are equal to 1. Let $\phi_m(\Delta t_{i+1}) = e^{-\kappa m^2 \pi^2 \Delta t_{i+1}}$ and then we may re-express the sum using matrix multiplication as

$$[\Phi(\Delta t_{i+1}) \tilde{x}(t_i)](\rho) = \boldsymbol{\alpha}^T(t_i) \tilde{\Phi}(\Delta t_{i+1}) \boldsymbol{\beta}(\rho) \quad (4.110)$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are as previously defined in Equations (4.35) and (4.36), respectively, and the N by N diagonal *state transition matrix* is defined as¹⁵

$$\tilde{\Phi}(\Delta t_{i+1}) \triangleq \begin{bmatrix} \phi_0(\Delta t_{i+1}) & 0 & \dots & 0 \\ 0 & \phi_1(\Delta t_{i+1}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \phi_{N-1}(\Delta t_{i+1}) \end{bmatrix} \quad (4.111)$$

$\tilde{\Phi}$ was set with bold type to emphasize that it is a matrix quantity and the tilde was added above to indicate that it is a finite-dimensional representation (or approximation) of the state transition operator Φ . An equivalent way of expressing the state transition matrix is by giving the diagonal entries

$$\left[\tilde{\Phi} \right]_n(t_{i+1} - t_i) = e^{-\kappa n^2 \pi^2 (t_{i+1} - t_i)}, n = 0, 1, \dots, N-1 \quad (4.112)$$

Additionally, we note that Equation (4.110) is a weighted inner product

$$[\Phi(\Delta t_{i+1}) \tilde{x}(t_i)](\rho) = \langle \boldsymbol{\alpha}(t_i), \boldsymbol{\beta}(\rho) \rangle_{\tilde{\Phi}(\Delta t_{i+1})} \quad (4.113)$$

¹⁵As a point of interest, we note that, per Theorem 6.4.4 of Naylor and Sell [154], there exists a basis that will yield a diagonal matrix representation for self-adjoint linear operators that map finite-dimensional Hilbert spaces to themselves...thus we can plainly see that our choice of bases used to express our state function has an actual impact on our matrix representations.

whereas $\tilde{x}(t_i) = [\mathcal{P} x(t_i)](\rho) = \boldsymbol{\alpha}^T(t_i) \boldsymbol{\beta}(\rho)$ is the usual inner product in \mathbb{R}^N , which can be thought of as an inner product with an identity for weighting. Note that, for zero propagation time, $\Delta t_{i+1} = 0$, $\tilde{\Phi}(\Delta t_{i+1})$ defined in Equation (4.111) is an identity matrix as anticipated!

4.3.6 The Equivalent Discrete-Time Input Distributor Matrix $\tilde{\mathbf{B}}_d(t_i)$. We have already determined the first term on the right-hand side of Equation (4.23), $\Phi(t_{i+1} - t_i) \tilde{x}(t_i)$, now we shall address the second term: $B_d(t_i) \tilde{u}(t_i)$, where

$$B_d(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1} - s) B ds \quad (4.17)$$

Before we employ any approximations, we have

$$B_d(t_i) u(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1} - s) B ds u(t_i) \quad (4.114)$$

which can be written as

$$B_d(t_i) u(t_i) = B \int_{t_i}^{t_{i+1}} \Phi(t_{i+1} - s) u(t_i) ds \quad (4.115)$$

since B is a constant and $u(t_i)$, which is piece-wise constant, is not a function of the integration variable s . Let

$$u(t_i, \rho) = \sum_{m=0}^{\infty} \nu_m(t_i) \beta_m(\rho) \quad (4.116)$$

and then

$$\tilde{u}(t_i, \rho) = \mathcal{P} u(t_i, \rho) = \sum_{m=0}^{N-1} \nu_m(t_i) \beta_m(\rho) \quad (4.117)$$

Hence by Equation (4.109)

$$\Phi(t_{i+1} - s) \tilde{u}(t_i) = \sum_{m=0}^{N-1} e^{-\kappa m^2 \pi^2 (t_{i+1} - s)} \nu_m(t_i) \beta_m(\rho) \quad (4.118)$$

Therefore, Equation (4.115) for $\tilde{u}(t_i)$ becomes

$$B_d(t_i) \tilde{u}(t_i) = B \int_{t_i}^{t_{i+1}} \sum_{m=0}^{N-1} e^{-\kappa m^2 \pi^2 (t_{i+1}-s)} \nu_m(t_i) \beta_m(\rho) ds \quad (4.119)$$

$$= B \sum_{m=0}^{N-1} \nu_m(t_i) \int_{t_i}^{t_{i+1}} e^{-\kappa m^2 \pi^2 (t_{i+1}-s)} ds \beta_m(\rho) \quad (4.120)$$

For the $m = 0$ case, the integrand is one, and thus the integral yields $[t_{i+1} - t_i]$, while for $m > 0$, integrating yields

$$\int_{t_i}^{t_{i+1}} e^{-\kappa m^2 \pi^2 (t_{i+1}-s)} ds = \frac{e^{-\kappa m^2 \pi^2 t_{i+1}}}{\kappa m^2 \pi^2} \left(e^{\kappa m^2 \pi^2 t_{i+1}} - e^{\kappa m^2 \pi^2 t_i} \right) \quad (4.121)$$

$$= \frac{1}{\kappa m^2 \pi^2} \left(1 - e^{-\kappa m^2 \pi^2 [t_{i+1}-t_i]} \right) \quad (4.122)$$

Thus,

$$B_d(t_i) \tilde{u}(t_i) = B \left\{ \nu_0(t_i) [t_{i+1} - t_i] \beta_0(\rho) + \sum_{m=1}^{N-1} \nu_m(t_i) \frac{1}{\kappa m^2 \pi^2} \left(1 - e^{-\kappa m^2 \pi^2 [t_{i+1}-t_i]} \right) \beta_m(\rho) \right\} \quad (4.123)$$

Now let the *equivalent discrete-time input distributor matrix*, $\tilde{\mathbf{B}}_d(t_i) \in \mathbb{R}^{N \times N}$, be defined as a real diagonal matrix with diagonal entries of

$$\left[\tilde{\mathbf{B}}_d \right]_n(t_i) = \begin{cases} B[t_{i+1} - t_i], & n = 0 \\ \frac{B \left[1 - e^{-\kappa n^2 \pi^2 (t_{i+1}-t_i)} \right]}{\kappa n^2 \pi^2}, & n = 1, 2, \dots, N-1 \end{cases} \quad (4.124)$$

As such, $\tilde{\mathbf{B}}_d(t_i)$ represents the operator $B_d(t_i)$ in finite dimensions. Therefore, we can write Equation (4.123) as

$$B_d(t_i) \tilde{u}(t_i) = \boldsymbol{\nu}^T(t_i) \tilde{\mathbf{B}}_d(t_i) \boldsymbol{\beta}(\rho) = \left[\tilde{\mathbf{B}}_d(t_i) \boldsymbol{\nu}(t_i) \right]_{\mathbb{B}_N} \quad (4.125)$$

where

$$\boldsymbol{\nu}(t_i) \equiv \begin{bmatrix} \nu_0(t_i) & \nu_1(t_i) & \cdots & \nu_{N-1}(t_i) \end{bmatrix}^T \quad (4.126)$$

is a vector of the first N input function coefficients — see Equation (4.117).

4.3.7 Propagate the Finite-Dimensional State Estimate. The approximate state propagation conditional moment is $E[\tilde{\mathbf{x}}(t_{i+1})|\mathbf{Z}(t_i) = \mathbf{Z}_i]$. Since $E[\mathbf{x}(t_{i+1})|\mathbf{Z}(t_i) = \mathbf{Z}_i]$ equals $\hat{x}(t_{i+1}^-)$; therefore $E[\tilde{\mathbf{x}}(t_{i+1})|\mathbf{Z}(t_i) = \mathbf{Z}_i]$ is $\tilde{\hat{x}}(t_{i+1}^-)$. Thus,

$$\tilde{\hat{x}}(t_{i+1}^-) = \hat{\boldsymbol{\alpha}}^T(t_{i+1}^-) \boldsymbol{\beta}(\rho) = [\hat{\boldsymbol{\alpha}}(t_{i+1}^-)]_{\mathbb{B}_N} \quad (4.127)$$

using the previously updated state estimate $\tilde{\hat{x}}(t_i^+)$

$$[\tilde{\hat{x}}(t_{i+1}^-)](\rho) = [\Phi(t_{i+1} - t_i) \tilde{\hat{x}}(t_i^+) + B_d(t_i) \tilde{u}(t_i)](\rho) \quad (4.128)$$

$$= \hat{\boldsymbol{\alpha}}^T(t_i^+) \tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) \boldsymbol{\beta}(\rho) + \boldsymbol{\nu}^T(t_i) \tilde{\mathbf{B}}_d(t_i) \boldsymbol{\beta}(\rho) \quad (4.129)$$

$$= [\tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) \hat{\boldsymbol{\alpha}}(t_i^+) + \tilde{\mathbf{B}}_d(t_i) \boldsymbol{\nu}(t_i)]_{\mathbb{B}_N} \quad (4.130)$$

Since the basis vector $\boldsymbol{\beta}(\rho)$ is the same for all states, propagating the coefficients for the approximate state function is the same as propagating the state itself, that is

$$\boxed{\hat{\boldsymbol{\alpha}}(t_{i+1}^-) = \tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) \hat{\boldsymbol{\alpha}}(t_i^+) + \tilde{\mathbf{B}}_d(t_i) \boldsymbol{\nu}(t_i)} \quad (4.131)$$

where the state transition matrix, $\tilde{\boldsymbol{\Phi}}(\Delta t_{i+1})$, is defined in Equation (4.111) and the equivalent discrete-time input distributor matrix, $\tilde{\mathbf{B}}_d(t_i)$, is defined in Equation (4.124).

4.3.8 The Equivalent Discrete-Time Dynamics Noise Covariance Matrix $\tilde{\mathbf{Q}}_d(t_i)$. Per Equation (4.19), we know that the equivalent discrete-time dynamics noise covariance operator is

$$Q_d(t_i) = E[\mathbf{w}_d(t_i) \diamond \mathbf{w}_d(t_i)] \quad (4.132)$$

Since the state transition operator is self-adjoint, i.e., $\Phi^* = \Phi$, and Q is a real number, we can rewrite Equation (4.19) as

$$Q_d(t_i) = Q \int_{t_i}^{t_{i+1}} \Phi(t_{i+1} - s) \Phi(t_{i+1} - s) ds \quad (4.133)$$

where Q_d is an infinite-dimensional bounded linear operator that acts on the state¹⁶. From previous sections we have acquired the tools needed to find the matrix representation for the covariance operator Q_d when it is applied to a finite-dimensional state. Thus,

$$Q_d(t_i) \tilde{x}(t_i) = Q \int_{t_i}^{t_{i+1}} \Phi(t_{i+1} - s) \Phi(t_{i+1} - s) ds \tilde{x}(t_i) \quad (4.134)$$

$$= Q \int_{t_i}^{t_{i+1}} \Phi(t_{i+1} - s) \Phi(t_{i+1} - s) \tilde{x}(t_i) ds \quad (4.135)$$

where we have factored $\tilde{x}(t_i) = \mathcal{P} x(t_i)$ into the integral. We can simplify the integrand by noting that $\{\Phi(\tau) : \tau \geq 0\}$ forms a semi-group of operators and thus

$$[\Phi(t_{i+1} - s) \Phi(t_{i+1} - s) \tilde{x}(t_i)](\rho) = [\Phi(t_{i+1} - s + t_{i+1} - s) \tilde{x}(t_i)](\rho) \quad (4.136)$$

$$= \boldsymbol{\alpha}^T(t_i) \tilde{\Phi}(t_{i+1} - s + t_{i+1} - s) \boldsymbol{\beta}(\rho) \quad (4.137)$$

$$= \boldsymbol{\alpha}^T(t_i) \tilde{\Phi}(t_{i+1} - s) \tilde{\Phi}(t_{i+1} - s) \boldsymbol{\beta}(\rho) \quad (4.138)$$

$$= \boldsymbol{\alpha}^T(t_i) \tilde{\Phi}^2(t_{i+1} - s) \boldsymbol{\beta}(\rho) \quad (4.139)$$

where the first equality is due to the semi-group property, the second line is analogous to Equation (4.110), the third line follows from the definition of $\tilde{\Phi}$, and the last line

¹⁶In the finite dimensional case, the equivalent discrete-time covariance was given in the form of an n -by- n matrix \mathbf{Q}_d . According to matrix multiplication rules, an n -by- n matrix may only be pre-multiplied by a matrix (or a vector) with n rows or post-multiplied by a matrix (or a vector) with n columns; this corresponds to the same size as the state vector. Thus it should come as no surprise that in our infinite-dimensional case, this equivalent discrete-time covariance operator may, in fact, act on an infinite-dimensional state.

is by convention. Thus, we can simplify Equation (4.135) to

$$Q_d(t_i) \tilde{x}(t_i) = Q \int_{t_i}^{t_{i+1}} \boldsymbol{\alpha}^T(t_i) \tilde{\boldsymbol{\Phi}}^2(t_{i+1} - s) \boldsymbol{\beta}(\rho) ds \quad (4.140)$$

$$= Q \int_{t_i}^{t_{i+1}} \left[\sum_{n=0}^{N-1} \alpha_n(t_i) \phi_n^2(t_{i+1} - s) \beta_n(\rho) \right] ds \quad (4.141)$$

Furthermore, since the sum is finite, we can extract all non “ s ” terms out of the integral to get:

$$Q_d(t_i) \tilde{x}(t_i) = Q \sum_{n=0}^{N-1} \alpha_n(t_i) \int_{t_i}^{t_{i+1}} \phi_n^2(t_{i+1} - s) ds \beta_n(\rho) \quad (4.142)$$

$$= Q \sum_{n=0}^{N-1} \alpha_n(t_i) \int_{t_i}^{t_{i+1}} e^{-2\kappa n^2 \pi^2 (t_{i+1} - s)} ds \beta_n(\rho) \quad (4.143)$$

Note that, for $n = 0$, the integrand becomes one, whereas for the $n \neq 0$ case, an exponential remains. Integrating yields

$$Q_d(t_i) \tilde{x}(t_i) = Q \left\{ \alpha_0(t_i) [t_{i+1} - t_i] \beta_0(\rho) + \sum_{n=1}^{N-1} \alpha_n(t_i) \frac{e^{-2\kappa n^2 \pi^2 t_{i+1}}}{2\kappa n^2 \pi^2} \left(e^{2\kappa n^2 \pi^2 t_{i+1}} - e^{2\kappa n^2 \pi^2 t_i} \right) \beta_n(\rho) \right\} \quad (4.144)$$

and then multiplying out the exponential gives

$$Q_d(t_i) \tilde{x}(t_i) = Q \left\{ \alpha_0(t_i) [t_{i+1} - t_i] \beta_0(\rho) + \sum_{n=1}^{N-1} \alpha_n(t_i) \frac{1}{2\kappa n^2 \pi^2} \left[1 - e^{-2\kappa n^2 \pi^2 (t_{i+1} - t_i)} \right] \beta_n(\rho) \right\} \quad (4.145)$$

Now let the *equivalent discrete-time dynamics noise covariance matrix*, $\tilde{\mathbf{Q}}_d(t_i) \in \mathbb{R}^{N \times N}$, be defined as a diagonal matrix with entries

$$\left[\tilde{\mathbf{Q}}_d\right]_n(t_i) = \begin{cases} Q[t_{i+1} - t_i], & n = 0 \\ \frac{Q[1 - e^{-2\kappa n^2 \pi^2(t_{i+1} - t_i)}]}{2\kappa n^2 \pi^2}, & n = 1, 2, \dots, N-1 \end{cases} \quad (4.146)$$

As such, $\tilde{\mathbf{Q}}_d(t_i)$ is the finite-dimensional representation of the equivalent discrete-time dynamics noise covariance operator $Q_d(t_i)$. Therefore, we can write Equation (4.145) as

$$Q_d(t_i) \tilde{x}(t_i) = \boldsymbol{\alpha}^T(t_i) \tilde{\mathbf{Q}}_d(t_i) \boldsymbol{\beta}(\rho) \quad (4.147)$$

4.3.9 First and Second Order Statistical Moments for the State Coefficients.

The discrete-time dynamics model for the approximate state function given by Equation (4.23) and rewritten here with decremented time indices as

$$\tilde{x}(t_i) = \Phi(t_i - t_{i-1}) \tilde{x}(t_{i-1}) + B_d(t_{i-1}) \tilde{u}(t_{i-1}) + \tilde{\mathbf{w}}_d(t_{i-1}) \quad (4.148)$$

gives us a model for propagating the state on a finite-dimensional subspace from one time sample to the next. Since we are, in effect, propagating a Gaussian conditional probability density function (PDF), we need to have knowledge of the first two statistical moments. The conditional state mean, corresponding to the propagation cycle, is

$$\begin{aligned} E[\tilde{x}(t_i^-) | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] \\ = E[\Phi(t_i - t_{i-1}) \tilde{x}(t_{i-1}) + B_d(t_{i-1}) \tilde{u}(t_{i-1}) + \tilde{\mathbf{w}}_d(t_{i-1}) | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] \end{aligned} \quad (4.149)$$

The Kalman filter computes two different conditional¹⁷ error covariance operators, $P(t_i^-)$ and $P(t_i^+)$, as defined in Definition 89, on page 3-83, by Equations (3.251) and (3.253) respectively. Without loss in generality, we will treat the conditional error covariance operator generically in the following development. Note that all error covariances are assumed to be taken using a conditional expectation operator; however, the conditioning will normally be suppressed to ease notation. Next, we use the definition of the error covariance operator to obtain

$$P \tilde{y} = E\{\tilde{\mathbf{e}} \diamond \tilde{\mathbf{e}}\} \tilde{y} \quad (4.150)$$

$$= E\{[\tilde{\mathbf{e}} \diamond \tilde{\mathbf{e}}] \tilde{y}\} \quad (4.151)$$

$$= E\{\tilde{\mathbf{e}} \langle \tilde{\mathbf{e}}, \tilde{y} \rangle\} \quad (4.152)$$

where we have used the definition of the outer product for functions, as defined in Equation (3.6) on page 3-10, to express the outer product in terms of an inner product as shown. Now if the state estimation error¹⁸ \mathbf{e} is approximated by $\tilde{\mathbf{e}} = \boldsymbol{\epsilon}^T \boldsymbol{\beta}$ and y is approximated with $\tilde{y} = \boldsymbol{\gamma}^T \boldsymbol{\beta}$, then expanding using summation notation yields

$$P \tilde{y} = E \left\{ \sum_{n=0}^{N-1} \epsilon_n \beta_n \left\langle \sum_{m=0}^{N-1} \epsilon_m \beta_m, \sum_{l=0}^{N-1} \gamma_l \beta_l \right\rangle \right\} \quad (4.153)$$

$$= E \left\{ \sum_{n=0}^{N-1} \epsilon_n \beta_n \sum_{m=0}^{N-1} \sum_{l=0}^{N-1} \epsilon_m \gamma_l \langle \beta_m, \beta_l \rangle \right\} \quad (4.154)$$

$$= E \left\{ \sum_{n=0}^{N-1} \epsilon_n \beta_n \sum_{m=0}^{N-1} \sum_{l=0}^{N-1} \epsilon_m \gamma_l \delta_{ml} \right\} \quad (4.155)$$

¹⁷Recall that the expectation operator E is conditioned on the measurement history, $\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}$, for the propagate cycle from time t_{i-1} to time t_i ; while E is conditioned on the measurement history, $\mathbf{Z}(t_i) = \mathbf{Z}_i$, for the state measurement update at time t_i .

¹⁸For ease of notation, the conditional nature of the expectation operator, as well as the dependencies on time t and space ρ for the state and error functions are suppressed in this development.

Then evaluating the Kronecker delta gives

$$P \tilde{y} = E \left\{ \sum_{n=0}^{N-1} \epsilon_n \beta_n \sum_{m=0}^{N-1} \epsilon_m \gamma_m \right\} \quad (4.156)$$

$$= E \left\{ \sum_{m=0}^{N-1} \gamma_m \epsilon_m \sum_{n=0}^{N-1} \epsilon_n \beta_n \right\} \quad (4.157)$$

$$= E \{ \boldsymbol{\gamma}^T \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \boldsymbol{\beta} \} \quad (4.158)$$

where the third line employs the compact vector notation.

Pulling the $\boldsymbol{\gamma}$ terms outside of the expectation operator gives

$$P \tilde{y} = \boldsymbol{\gamma}^T E \{ \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \} \boldsymbol{\beta} \quad (4.159)$$

$$= \boldsymbol{\gamma}^T \tilde{\mathbf{P}} \boldsymbol{\beta} \quad (4.160)$$

where the conditional error covariance operator P has a finite-dimensional matrix representation of

$$\tilde{\mathbf{P}} \triangleq E \{ \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \} \quad (4.161)$$

For completeness, we give the full notation for the propagation cycle

$$\tilde{\mathbf{P}}(t_i^-) \triangleq E \{ \boldsymbol{\epsilon}(t_i^-) \boldsymbol{\epsilon}^T(t_i^-) | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1} \} \quad (4.162)$$

and the update step

$$\tilde{\mathbf{P}}(t_i^+) \triangleq E \{ \boldsymbol{\epsilon}(t_i^+) \boldsymbol{\epsilon}^T(t_i^+) | \mathbf{Z}(t_i) = \mathbf{Z}_i \} \quad (4.163)$$

4.3.10 The Propagation Error Covariance Matrix $\tilde{\mathbf{P}}(t_{i+1}^-)$. The propagation error covariance operator equation, given by Equation (3.271) of Theorem 91,

at time t_{i+1} holds for all states $x \in \mathbb{L}_{[0,1]}^2$

$$P(t_{i+1}^-) x(t_i) = [\Phi(\Delta t_{i+1}) P(t_i^+) \Phi^*(\Delta t_{i+1}) + Q_d(t_i)] x(t_i) \quad (4.164)$$

$$= [\Phi(\Delta t_{i+1}) E\{\mathbf{e}(t_i^+) \diamond \mathbf{e}(t_i^+)\} \Phi^*(\Delta t_{i+1}) + Q_d(t_i)] x(t_i) \quad (4.165)$$

where the update error covariance operator is $P(t_i^+) = E\{\mathbf{e}(t_i^+) \diamond \mathbf{e}(t_i^+)\}$. For a finite-dimensional approximation of the state function, we have

$$P(t_{i+1}^-) \tilde{x}(t_i) = [\Phi(\Delta t_{i+1}) E\{\tilde{\mathbf{e}}(t_i^+) \diamond \tilde{\mathbf{e}}(t_i^+)\} \Phi^*(\Delta t_{i+1}) + Q_d(t_i)] \tilde{x}(t_i) \quad (4.166)$$

$$= \Phi(\Delta t_{i+1}) E\{\tilde{\mathbf{e}}(t_i^+) \diamond \tilde{\mathbf{e}}(t_i^+)\} \Phi^*(\Delta t_{i+1}) \tilde{x}(t_i) + Q_d(t_i) \tilde{x}(t_i) \quad (4.167)$$

Since we already know $Q_d(t_i) \tilde{x}(t_i)$ from Equation (4.147), we shall address the first term in Equation (4.167)

$$\begin{aligned} & \Phi(\Delta t_{i+1}) E\{\tilde{\mathbf{e}}(t_i^+) \diamond \tilde{\mathbf{e}}(t_i^+)\} \Phi^*(\Delta t_{i+1}) \tilde{x}(t_i) \\ &= \Phi(\Delta t_{i+1}) E\{[\tilde{\mathbf{e}}(t_i^+) \diamond \tilde{\mathbf{e}}(t_i^+)] \Phi^*(\Delta t_{i+1}) \tilde{x}(t_i)\} \end{aligned} \quad (4.168)$$

Applying the definition of the outer product from Equation (3.6) yields

$$\begin{aligned} & \Phi(\Delta t_{i+1}) E\{\tilde{\mathbf{e}}(t_i^+) \diamond \tilde{\mathbf{e}}(t_i^+)\} \Phi^*(\Delta t_{i+1}) \tilde{x}(t_i) \\ &= \Phi(\Delta t_{i+1}) E\{\tilde{\mathbf{e}}(t_i^+) \langle \tilde{\mathbf{e}}(t_i^+), \Phi^*(\Delta t_{i+1}) \tilde{x}(t_i) \rangle\} \end{aligned} \quad (4.169)$$

Expanding $\tilde{\mathbf{e}}(t_i^+) \langle \tilde{\mathbf{e}}(t_i^+), \Phi^*(\Delta t_{i+1}) \tilde{x}(t_i) \rangle$ using summation notations yields

$$\begin{aligned} & \tilde{\mathbf{e}}(t_i^+) \langle \tilde{\mathbf{e}}(t_i^+), \Phi^*(\Delta t_{i+1}) \tilde{x}(t_i) \rangle \\ &= \sum_{n=0}^{N-1} \epsilon_n(t_i^+) \beta_n(\rho) \left\langle \sum_{m=0}^{N-1} \epsilon_m(t_i^+) \beta_m(\rho), \sum_{l=0}^{N-1} \phi_l(\Delta t_{i+1}) \alpha_l(t_i) \beta_l(\rho) \right\rangle \end{aligned} \quad (4.170)$$

and then moving the inner product inside (since all the other terms are independent of ρ) produces

$$\begin{aligned} & \tilde{\mathbf{e}}(t_i^+) \langle \tilde{\mathbf{e}}(t_i^+), \Phi^*(\Delta t_{i+1}) \tilde{\mathbf{x}}(t_i) \rangle \\ &= \sum_{n=0}^{N-1} \epsilon_n(t_i^+) \beta_n(\rho) \sum_{m=0}^{N-1} \sum_{l=0}^{N-1} \epsilon_m(t_i^+) \phi_l(\Delta t_{i+1}) \alpha_l(t_i) \langle \beta_m(\rho), \beta_l(\rho) \rangle \end{aligned} \quad (4.171)$$

$$= \sum_{n=0}^{N-1} \epsilon_n(t_i^+) \beta_n(\rho) \sum_{m=0}^{N-1} \sum_{l=0}^{N-1} \epsilon_m(t_i^+) \phi_l(\Delta t_{i+1}) \alpha_l(t_i) \delta_{ml}(\rho) \quad (4.172)$$

$$= \sum_{n=0}^{N-1} \epsilon_n(t_i^+) \beta_n(\rho) \sum_{l=0}^{N-1} \epsilon_l(t_i^+) \phi_l(\Delta t_{i+1}) \alpha_l(t_i) \quad (4.173)$$

where we recognized and then evaluated the Kronecker delta $\delta_{ml} = \langle \beta_m, \beta_l \rangle$.

Reordering terms and then writing in terms of vector notation yields

$$\begin{aligned} & \tilde{\mathbf{e}}(t_i^+) \langle \tilde{\mathbf{e}}(t_i^+), \Phi^*(\Delta t_{i+1}) \tilde{\mathbf{x}}(t_i) \rangle \\ &= \sum_{l=0}^{N-1} \phi_l(\Delta t_{i+1}) \alpha_l(t_i) \epsilon_l(t_i^+) \sum_{n=0}^{N-1} \epsilon_n(t_i^+) \beta_n(\rho) \end{aligned} \quad (4.174)$$

$$= \boldsymbol{\alpha}^T(t_i) \tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) \boldsymbol{\epsilon}(t_i^+) \boldsymbol{\epsilon}^T(t_i^+) \boldsymbol{\beta}(\rho) \quad (4.175)$$

Now we re-apply the transition operator $\Phi(\Delta t_{i+1})$ and expectation operator¹⁹ E in order to obtain the first term of Equation (4.167)

$$\begin{aligned} & \Phi(\Delta t_{i+1}) E \left\{ \boldsymbol{\alpha}^T(t_i) \tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) \boldsymbol{\epsilon}(t_i^+) \boldsymbol{\epsilon}^T(t_i^+) \boldsymbol{\beta}(\rho) \right\} \\ &= \Phi(\Delta t_{i+1}) \left[\boldsymbol{\alpha}^T(t_i) \tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) E \left\{ \boldsymbol{\epsilon}(t_i^+) \boldsymbol{\epsilon}^T(t_i^+) \right\} \boldsymbol{\beta}(\rho) \right] \end{aligned} \quad (4.176)$$

$$= \boldsymbol{\alpha}^T(t_i) \tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) E \left\{ \boldsymbol{\epsilon}(t_i^+) \boldsymbol{\epsilon}^T(t_i^+) \right\} \tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) \boldsymbol{\beta}(\rho) \quad (4.177)$$

$$= \boldsymbol{\alpha}^T(t_i) \tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) \tilde{\mathbf{P}}(t_i^+) \tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) \boldsymbol{\beta}(\rho) \quad (4.178)$$

where as before $\tilde{\mathbf{P}}(t_i^+) = E \left\{ \boldsymbol{\epsilon}(t_i^+) \boldsymbol{\epsilon}^T(t_i^+) \right\}$ and we see that the state transition operator $\tilde{\boldsymbol{\Phi}}(\Delta t_{i+1})$ merely weights the inner product $\boldsymbol{\alpha}^T(t_i) \tilde{\mathbf{P}}(t_i^+) \tilde{\boldsymbol{\Phi}}(\Delta t_{i+1}) \boldsymbol{\beta}(\rho)$ with

¹⁹This is still the appropriately conditioned expectation operator.

the diagonal matrix $\tilde{\Phi}(\Delta t_{i+1})$. Using Equations (4.147) and (4.178) we get

$$\begin{aligned} P(t_{i+1}^-) \tilde{x}(t_i) \\ = \alpha^T(t_i) \tilde{\Phi}(\Delta t_{i+1}) \tilde{\mathbf{P}}(t_i^+) \tilde{\Phi}(\Delta t_{i+1}) \beta(\rho) + \alpha^T(t_i) \tilde{\mathbf{Q}}_d(t_i) \beta(\rho) \end{aligned} \quad (4.179)$$

$$= \alpha^T(t_i) \left[\tilde{\Phi}(\Delta t_{i+1}) \tilde{\mathbf{P}}(t_i^+) \tilde{\Phi}(\Delta t_{i+1}) + \tilde{\mathbf{Q}}_d(t_i) \right] \beta(\rho) \quad (4.180)$$

Since Equation (4.180) holds for any state, $x(t_i)$, the *error covariance matrix for propagation* is

$$\boxed{\tilde{\mathbf{P}}(t_{i+1}^-) = \tilde{\Phi}(\Delta t_{i+1}) \tilde{\mathbf{P}}(t_i^+) \tilde{\Phi}(\Delta t_{i+1}) + \tilde{\mathbf{Q}}_d(t_i)} \quad (4.181)$$

where the state transition matrix, $\tilde{\Phi}(\Delta t_{i+1})$, is defined in Equation (4.111) and the equivalent discrete-time diffusion matrix, $\tilde{\mathbf{Q}}_d(t_i)$, is defined in Equation (4.146). Thus, propagating the approximate finite-dimensional error covariance operator, $\tilde{\mathbf{P}}(t_{i+1}^-)$, is achieved by a weighted inner product that depends on a diagonal matrix, $\tilde{\Phi}(\Delta t_{i+1})$, related to the state transition operator Φ ; a diagonal matrix $\tilde{\mathbf{Q}}_d(t_i)$, related to the equivalent discrete-time diffusion operator, Q_d ; and a symmetric matrix, $\tilde{\mathbf{P}}(t_i^+)$, due to the conditional error covariance following the measurement update.

4.3.11 The Measurement Distributor Matrix $\tilde{\mathbf{H}}$. According to Theorem 3.11 of Hoffman and Kunze [85], there exists a unique matrix representation for a finite-dimensional transformation relative to some particular ordered basis. So our first task is to create a finite-dimensional transformation. The measurement distributor transformation, \mathbf{H} , as defined for an arbitrary realization of the state function, x , is

$$\mathbf{H} x(t_i) = \begin{bmatrix} \int_0^{1/M} x(t_i, \rho) d\rho \\ \vdots \\ \int_{(M-1)/M}^1 x(t_i, \rho) d\rho \end{bmatrix} \quad (4.182)$$

Next, \mathbf{H} applied to a finite-dimensional approximation of the state, $\tilde{x}(t_i)$, is

$$\mathbf{H} \tilde{x}(t_i) = \begin{bmatrix} \int_0^{1/M} \tilde{x}(t_i, \rho) d\rho \\ \vdots \\ \int_{(M-1)/M}^1 \tilde{x}(t_i, \rho) d\rho \end{bmatrix} \quad (4.183)$$

Using a truncated Fourier series expression, $\tilde{x}(t_i, \rho) = \boldsymbol{\alpha}^T(t_i) \boldsymbol{\beta}(\rho)$ we get

$$\mathbf{H} \tilde{x}(t_i) = \begin{bmatrix} \int_0^{1/M} \boldsymbol{\alpha}^T(t_i) \boldsymbol{\beta}(\rho) d\rho \\ \vdots \\ \int_{(M-1)/M}^1 \boldsymbol{\alpha}^T(t_i) \boldsymbol{\beta}(\rho) d\rho \end{bmatrix} \quad (4.184)$$

The m th element of the vector can be simplified using the *integrated basis* $\mu_{m,n}$ for $m = \{1, \dots, M\}$ and $n = \{0, \dots, N-1\}$

$$\mu_{m,n} \triangleq \int_{(m-1)/M}^{m/M} \beta_n(\rho) d\rho \quad (4.185)$$

as

$$\int_{(m-1)/M}^{m/M} \boldsymbol{\alpha}^T(t_i) \boldsymbol{\beta}(\rho) d\rho = \int_{(m-1)/M}^{m/M} \sum_{n=0}^{N-1} \alpha_n(t_i) \beta_n(\rho) d\rho \quad (4.186)$$

$$= \sum_{n=0}^{N-1} \alpha_n(t_i) \int_{(m-1)/M}^{m/M} \beta_n(\rho) d\rho \quad (4.187)$$

$$= \sum_{n=0}^{N-1} \alpha_n(t_i) \mu_{m,n} \quad (4.188)$$

$$= \boldsymbol{\mu}_m^T \boldsymbol{\alpha}(t_i) \quad (4.189)$$

where

$$\boldsymbol{\mu}_m \equiv \begin{bmatrix} \mu_{m,0} & \mu_{m,1} & \cdots & \mu_{m,N-1} \end{bmatrix}^T \quad (4.190)$$

Then we can write Equation (4.184) as

$$\mathbf{H} \tilde{x}(t_i) = \begin{bmatrix} \boldsymbol{\mu}_1^T \boldsymbol{\alpha}(t_i) \\ \vdots \\ \boldsymbol{\mu}_M^T \boldsymbol{\alpha}(t_i) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \vdots \\ \boldsymbol{\mu}_M^T \end{bmatrix} \boldsymbol{\alpha}(t_i) \quad (4.191)$$

By placing the integrated basis elements $\{\boldsymbol{\mu}_m : 1 \leq m \leq M\}$ in a matrix we get the *measurement distributor matrix* defined by

$$\tilde{\mathbf{H}} \triangleq \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \vdots \\ \boldsymbol{\mu}_M^T \end{bmatrix} = \begin{bmatrix} \mu_{1,0} & \mu_{1,1} & \cdots & \mu_{1,N-1} \\ \vdots & \vdots & & \vdots \\ \mu_{M,0} & \mu_{M,1} & \cdots & \mu_{M,N-1} \end{bmatrix} \in \mathbb{R}^{M \times N} \quad (4.192)$$

Note that $\tilde{\mathbf{H}}$ is time-invariant and can be pre-computed analytically or with a numerical technique. Hence Equation (4.191) can be written as

$$\mathbf{H} \tilde{x}(t_i) = \tilde{\mathbf{H}} \boldsymbol{\alpha}(t_i) \quad (4.193)$$

For the problem at hand, Equation (4.185) can be written as

$$\mu_{m,n} = \begin{cases} \int_{(m-1)/M}^{m/M} d\rho, & n = 0 \\ \int_{(m-1)/M}^{m/M} \sqrt{2} \cos(n\pi\rho) d\rho, & n = 1, 2, \dots, N-1 \end{cases} \quad (4.194)$$

and then evaluated as

$$\mu_{m,n} = \begin{cases} \frac{1}{M}, & n = 0 \\ \frac{\sqrt{2}}{n\pi} \sin(n\pi\rho) \Big|_{(m-1)/M}^{m/M}, & n = 1, 2, \dots, N-1 \end{cases} \quad (4.195)$$

Note that while we may pre-compute and store this matrix since it is independent of time, we could compute it at every step if desired. Now if the basis \mathbb{B}_N or the measurement distributor transformation were a function of time, then the matrix $\tilde{\mathbf{H}}$ would have to be computed at each time step since the matrix would no longer be time-invariant. So, instead of propagating/updating the entire estimate of the state function approximation, the estimated coefficients may be propagated as $\hat{\boldsymbol{\alpha}}(t_i^-)$ and updated to $\hat{\boldsymbol{\alpha}}(t_i^+)$ using an online digital computer.

4.3.12 The Residual Covariance Matrix $\tilde{\mathbf{A}}(t_i)$. In Section 4.3.3 we began with $\mathbf{A}(t_i) = \mathbf{H} P(t_i^-) \mathbf{H}^* + \mathbf{R}(t_i)$ and ended with

$$\mathbf{A}(t_i) = \begin{bmatrix} E\{\eta_1(t_i^-) \eta_1(t_i^-)\} & \cdots & E\{\eta_1(t_i^-) \eta_M(t_i^-)\} \\ \vdots & \ddots & \vdots \\ E\{\eta_M(t_i^-) \eta_1(t_i^-)\} & \cdots & E\{\eta_M(t_i^-) \eta_M(t_i^-)\} \end{bmatrix} + \mathbf{R}(t_i) \quad (4.82)$$

where

$$E\{\eta_m(t_i^-) \eta_n(t_i^-)\} = \int_{(m-1)/M}^{m/M} \int_{(n-1)/M}^{n/M} E\{\mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho')\} d\rho' d\rho \quad (4.80)$$

In this section, we shall evaluate $E\{\eta_m(t_i^-) \eta_n(t_i^-)\}$ using the approximate error $\tilde{\mathbf{e}}(t_i^-, \rho)$ in order to calculate $\tilde{\mathbf{A}}(t_i)$. Note that, for the first time, we have not reduced the dimension, but rather, we have simply introduced an approximation so that we could implement the operator using a digital algorithm.

Substituting the approximate error in Equation (4.39) in for the error in Equation (4.80) yields

$$E\{\tilde{\eta}_m(t_i^-) \tilde{\eta}_n(t_i^-)\} = \int_{(m-1)/M}^{m/M} \int_{(n-1)/M}^{n/M} E\{\tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho')\} d\rho' d\rho \quad (4.196)$$

Now use Equation (4.39) to obtain

$$E\{\tilde{\eta}_m(t_i^-) \tilde{\eta}_n(t_i^-)\} = \int_{(m-1)/M}^{m/M} \int_{(n-1)/M}^{n/M} E\left\{\sum_{j=0}^{N-1} \epsilon_j(t_i^-) \beta_j(\rho) \sum_{k=0}^{N-1} \epsilon_k(t_i^-) \beta_k(\rho')\right\} d\rho' d\rho \quad (4.197)$$

$$= \int_{(m-1)/M}^{m/M} \int_{(n-1)/M}^{n/M} E\left\{\sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \epsilon_j(t_i^-) \beta_j(\rho) \epsilon_k(t_i^-) \beta_k(\rho')\right\} d\rho' d\rho \quad (4.198)$$

Then, moving the expectation operator inside the double sum results in

$$E\{\tilde{\eta}_m(t_i^-) \tilde{\eta}_n(t_i^-)\} = \int_{(m-1)/M}^{m/M} \int_{(n-1)/M}^{n/M} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} E\{\epsilon_j(t_i^-) \beta_j(\rho) \epsilon_k(t_i^-) \beta_k(\rho')\} d\rho' d\rho \quad (4.199)$$

$$= \int_{(m-1)/M}^{m/M} \int_{(n-1)/M}^{n/M} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} E\{\epsilon_j(t_i^-) \epsilon_k(t_i^-)\} \beta_j(\rho) \beta_k(\rho') d\rho' d\rho \quad (4.200)$$

Since the sums are finite, we can factor out the terms independent of ρ and ρ'

$$E\{\tilde{\eta}_m(t_i^-) \tilde{\eta}_n(t_i^-)\} = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} E\{\epsilon_j(t_i^-) \epsilon_k(t_i^-)\} \int_{(m-1)/M}^{m/M} \int_{(n-1)/M}^{n/M} \beta_j(\rho) \beta_k(\rho') d\rho' d\rho \quad (4.201)$$

$$= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} E\{\epsilon_j(t_i^-) \epsilon_k(t_i^-)\} \int_{(m-1)/M}^{m/M} \beta_j(\rho) d\rho \int_{(n-1)/M}^{n/M} \beta_k(\rho') d\rho' \quad (4.202)$$

$$= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} E\{\epsilon_j(t_i^-) \epsilon_k(t_i^-)\} \mu_{m,j} \mu_{n,k} \quad (4.203)$$

where the $\mu_{m,j}$ is the integrated basis that we defined in Equation (4.185). Using matrix notation, we can rewrite the m th element as

$$E\{\tilde{\eta}_m(t_i^-) \tilde{\eta}_n(t_i^-)\} = \boldsymbol{\mu}_m^T E\{\boldsymbol{\epsilon}(t_i^-) \boldsymbol{\epsilon}^T(t_i^-)\} \boldsymbol{\mu}_n = \boldsymbol{\mu}_m^T \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\mu}_n \quad (4.204)$$

Hence our approximation of $\mathbf{H} P(t_i^-) \mathbf{H}^*$ is

$$\begin{aligned}
& \begin{bmatrix} E\{\tilde{\eta}_1(t_i^-) \tilde{\eta}_1(t_i^-)\} & \cdots & E\{\tilde{\eta}_1(t_i^-) \tilde{\eta}_M(t_i^-)\} \\ \vdots & \ddots & \vdots \\ E\{\tilde{\eta}_M(t_i^-) \tilde{\eta}_1(t_i^-)\} & \cdots & E\{\tilde{\eta}_M(t_i^-) \tilde{\eta}_M(t_i^-)\} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\mu}_1^T \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\mu}_1 & \cdots & \boldsymbol{\mu}_1^T \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\mu}_M \\ \vdots & \ddots & \vdots \\ \boldsymbol{\mu}_M^T \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\mu}_1 & \cdots & \boldsymbol{\mu}_M^T \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\mu}_M \end{bmatrix} \quad (4.205)
\end{aligned}$$

$$= \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \vdots \\ \boldsymbol{\mu}_M^T \end{bmatrix} \tilde{\mathbf{P}}(t_i^-) \begin{bmatrix} \boldsymbol{\mu}_1 & \cdots & \boldsymbol{\mu}_M \end{bmatrix} \quad (4.206)$$

$$= \tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \quad (4.207)$$

Therefore, the approximate *filter-computed residual covariance matrix* at time t_i is defined as

$$\tilde{\mathbf{A}}(t_i) = \tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \in \mathbb{R}^{M \times M} \quad (4.208)$$

4.3.13 The Kalman Gain Transformation Matrix $\tilde{\mathbf{K}}$. In Section 4.3.4 we found that the gain $K(t_i) = P(t_i^-) H^* \mathbf{A}^{-1}(t_i)$ can be written as

$$K(t_i) = \begin{bmatrix} \int_0^{1/M} E \{ \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') \} d\rho' \\ \vdots \\ \int_{(M-1)/M}^1 E \{ \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') \} d\rho' \end{bmatrix}^T \mathbf{A}^{-1}(t_i) \quad (4.95)$$

In order to evaluate the Kalman gain, we shall use the approximate error function, $\tilde{\mathbf{e}}(t_i^-, \rho) = \sum_{j=0}^{N-1} \epsilon_j(t_i^-) \beta_j(\rho)$ and the approximate filter-computed error covariance, $\mathbf{A}(t_i)$ given in Equation (4.208).

Substituting in for the approximate error function, we obtain the m th element of the row vector in Equation (4.95)

$$\begin{aligned} & \int_{(m-1)/M}^{m/M} E \left\{ \tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho') \right\} d\rho' \\ &= \int_{(m-1)/M}^{m/M} E \left\{ \sum_{j=0}^{N-1} \epsilon_j(t_i^-) \beta_j(\rho) \sum_{k=0}^{N-1} \epsilon_k(t_i^-) \beta_k(\rho') \right\} d\rho' \end{aligned} \quad (4.209)$$

$$= \int_{(m-1)/M}^{m/M} E \left\{ \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \epsilon_j(t_i^-) \beta_j(\rho) \epsilon_k(t_i^-) \beta_k(\rho') \right\} d\rho' \quad (4.210)$$

$$= \int_{(m-1)/M}^{m/M} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} E \left\{ \epsilon_j(t_i^-) \beta_j(\rho) \epsilon_k(t_i^-) \beta_k(\rho') \right\} d\rho' \quad (4.211)$$

$$= \int_{(m-1)/M}^{m/M} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} E \left\{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \right\} \beta_j(\rho) \beta_k(\rho') d\rho' \quad (4.212)$$

Noting that the integral of a finite sum is a sum of integrals

$$\begin{aligned} & \int_{(m-1)/M}^{m/M} E \left\{ \tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho') \right\} d\rho' \\ &= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} E \left\{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \right\} \beta_j(\rho) \int_{(m-1)/M}^{m/M} \beta_k(\rho') d\rho' \end{aligned} \quad (4.213)$$

$$= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} E \left\{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \right\} \beta_j(\rho) \mu_{m,k} \quad (4.214)$$

where the integrated basis elements $\mu_{m,k}$ was defined in Equation (4.185). Now using vector notation, we can write Equation (4.214) as

$$\int_{(m-1)/M}^{m/M} E \left\{ \tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho') \right\} d\rho' = \boldsymbol{\beta}^T(\rho) \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\mu}_m \quad (4.215)$$

Collecting the m th element from Equation (4.215) and the results from Equation (4.208) for the projected filter-computed residual covariance, we have the representation of the Kalman gain transformation in the subspace chosen, also known as the

Kalman gain matrix:

$$\tilde{\mathbf{K}}(t_i) = \left[\boldsymbol{\beta}^T(\rho) \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\mu}_1 \cdots \boldsymbol{\beta}^T(\rho) \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\mu}_M \right] \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \quad (4.216)$$

$$= \boldsymbol{\beta}^T(\rho) \tilde{\mathbf{P}}(t_i^-) \begin{bmatrix} \boldsymbol{\mu}_1 & \cdots & \boldsymbol{\mu}_M \end{bmatrix} \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \quad (4.217)$$

$$= \boldsymbol{\beta}^T(\rho) \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \quad (4.218)$$

Thus the Kalman gain matrix with respect to the basis \mathbb{B}_N is²⁰

$$\tilde{\mathbf{K}}(t_i) = \left[\tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \right]_{\mathbb{B}_N} \in \mathbb{R}^{N \times M} \quad (4.219)$$

4.3.14 *The Updated State Estimate $\hat{x}(t_i^+)$.* The approximate updated state estimate is $\tilde{x}(t_i^+)$

$$\tilde{x}(t_i^+) = \tilde{x}(t_i^-) + \tilde{\mathbf{K}}(t_i) \tilde{\mathbf{r}}(t_i) \quad (4.220)$$

$$= \hat{\boldsymbol{\alpha}}^T(t_i^-) \boldsymbol{\beta}(\rho) + \tilde{\mathbf{K}}(t_i) \tilde{\mathbf{r}}(t_i) \quad (4.221)$$

where the approximate measurement residual — approximate because the predicted measurement is an approximation — is given by

$$\tilde{\mathbf{r}}(t_i) = \mathbf{z}_i - \mathbf{H} \tilde{x}(t_i^-) = \mathbf{z}_i - \tilde{\mathbf{H}} \hat{\boldsymbol{\alpha}}(t_i^-) \quad (4.222)$$

Substituting in for the Kalman gain matrix defined in Equation (4.218) yields

$$\tilde{x}(t_i^+) = \hat{\boldsymbol{\alpha}}^T(t_i^-) \boldsymbol{\beta}(\rho) + \boldsymbol{\beta}^T(\rho) \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{r}}(t_i) \quad (4.223)$$

$$= \hat{\boldsymbol{\alpha}}^T(t_i^-) \boldsymbol{\beta}(\rho) + \left\{ \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{r}}(t_i) \right\}^T \boldsymbol{\beta}(\rho) \quad (4.224)$$

$$= \left[\hat{\boldsymbol{\alpha}}(t_i^-) + \left\{ \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{r}}(t_i) \right\} \right]^T \boldsymbol{\beta}(\rho) \quad (4.225)$$

²⁰See the coordinate vector notation introduced in Equation (4.37) on page 4-15.

where the second equality follows from the fact that the transpose of a scalar is a scalar. Additionally, since the basis \mathbb{B}_N is the same for all functions, we can write the approximate state update in terms of the basis

$$\tilde{x}(t_i^+) = \left[\hat{\alpha}(t_i^-) + \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{r}}(t_i) \right]_{\mathbb{B}_N} \quad (4.226)$$

Thus we can update the state by *updating the approximate state function coefficients*

$$\boxed{\hat{\alpha}(t_i^+) = \hat{\alpha}(t_i^-) + \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{r}}(t_i)} \quad (4.227)$$

where the propagated state coefficients, $\hat{\alpha}(t_i^-)$, are defined in Equation (4.131), the propagated error covariance matrix, $\tilde{\mathbf{P}}(t_i^-)$, is defined in Equation (4.162) and computed in Equation (4.181), the measurement distributor matrix, $\tilde{\mathbf{H}}$, is described in Equation (4.192), and the measurement noise covariance, $\mathbf{R}(t_i)$, is assumed known.

4.3.15 The Updated Error Covariance Operator $P(t_i^+)$. To generate a matrix representation for the updated error covariance operator we begin by recasting Equation (3.269) for this problem as

$$P(t_i^+) = P(t_i^-) - K(t_i) \mathbf{H} P(t_i^-) \quad (4.228)$$

for an arbitrary state x at time t_i

$$P(t_i^+) x(t_i) = [P(t_i^-) - K(t_i) \mathbf{H} P(t_i^-)] x(t_i) \quad (4.229)$$

$$= P(t_i^-) x(t_i) - K(t_i) \mathbf{H} P(t_i^-) x(t_i) \quad (4.230)$$

By definition,

$$P(t_i^-) x(t_i) = E\{[\mathbf{e}(t_i^-) \diamond \mathbf{e}(t_i^-)] | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\} x(t_i) \quad (4.231)$$

and (dropping the explicit conditioning on the measurement history) this can be written as

$$P(t_i^-) x(t_i) = E\{[\mathbf{e}(t_i^-) \diamond \mathbf{e}(t_i^-)]x(t_i)\} \quad (4.232)$$

$$= E\{\mathbf{e}(t_i^-) \langle \mathbf{e}(t_i^-), x(t_i) \rangle_{\mathbb{L}^2}\} \quad (4.233)$$

where the second line follows from the definition of the outer product given in Equation (3.6) on page 3-10. Substituting this result into Equation (4.230) and then expanding the measurement distributor transformation in the following line yields

$$P(t_i^+) x(t_i) = E\{\mathbf{e}(t_i^-) \langle \mathbf{e}(t_i^-), x(t_i) \rangle_{\mathbb{L}^2}\} - K(t_i) \mathbf{H} E\{\mathbf{e}(t_i^-) \langle \mathbf{e}(t_i^-), x(t_i) \rangle_{\mathbb{L}^2}\} \quad (4.234)$$

Explicitly writing out the spatial dependence of the state and the error yields

$$P(t_i^+) x(t_i) = E\{\mathbf{e}(t_i^-, \rho) \langle \mathbf{e}(t_i^-, \rho), x(t_i, \rho) \rangle_{\mathbb{L}^2}\} - K(t_i) \begin{bmatrix} \int_0^{1/M} E\{\mathbf{e}(t_i^-, \rho) \langle \mathbf{e}(t_i^-, \rho), x(t_i, \rho) \rangle_{\mathbb{L}^2}\} d\rho \\ \vdots \\ \int_{(M-1)/M}^1 E\{\mathbf{e}(t_i^-, \rho) \langle \mathbf{e}(t_i^-, \rho), x(t_i, \rho) \rangle_{\mathbb{L}^2}\} d\rho \end{bmatrix} \quad (4.235)$$

The first term in Equation (4.235) and the repeated integrand are the same; evaluating yields

$$E\{\mathbf{e}(t_i^-, \rho) \langle \mathbf{e}(t_i^-, \rho), x(t_i, \rho) \rangle_{\mathbb{L}^2}\} = E\left\{\mathbf{e}(t_i^-, \rho) \int_0^1 \mathbf{e}(t_i^-, \rho') x(t_i, \rho') d\rho'\right\} \quad (4.236)$$

$$= E\left\{\int_0^1 \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') x(t_i, \rho') d\rho'\right\} \quad (4.237)$$

$$= \int_0^1 E\{\mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') x(t_i, \rho')\} d\rho' \quad (4.238)$$

$$= \int_0^1 E\{\mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho')\} x(t_i, \rho') d\rho' \quad (4.239)$$

Substituting the results of Equation (4.239) back into Equation (4.235) gives

$$P(t_i^+) x(t_i) = P(t_i^-) x(t_i) - K(t_i) \mathbf{H} P(t_i^-) x(t_i) \quad (4.240)$$

$$= \int_0^1 E \{ \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') \} x(t_i, \rho') d\rho' \\ - K(t_i) \begin{bmatrix} \int_0^{1/M} \int_0^1 E \{ \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') \} x(t_i, \rho') d\rho' d\rho \\ \vdots \\ \int_{(M-1)/M}^1 \int_0^1 E \{ \mathbf{e}(t_i^-, \rho) \mathbf{e}(t_i^-, \rho') \} x(t_i, \rho') d\rho' d\rho \end{bmatrix} \quad (4.241)$$

where the Kalman gain is as defined in Equation (4.95).

Now that we have written out $P(t_i^+) x(t_i)$, we can substitute in the finite-dimensional approximations to determine the matrix representation for $P(t_i^+)$. As before, we replace the infinite-dimensional error and state functions with functions via the truncated Fourier series as and the Kalman gain approximation was given by Equation (4.219), thus

$$P(t_i^+) \tilde{x}(t_i) = P(t_i^-) \tilde{x}(t_i) - \tilde{\mathbf{K}}(t_i) \mathbf{H} P(t_i^-) \tilde{x}(t_i) \quad (4.242)$$

$$= \int_0^1 E \{ \tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho') \} \tilde{x}(t_i, \rho') d\rho' \\ - \tilde{\mathbf{K}}(t_i) \begin{bmatrix} \int_0^{1/M} \int_0^1 E \{ \tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho') \} \tilde{x}(t_i, \rho') d\rho' d\rho \\ \vdots \\ \int_{(M-1)/M}^1 \int_0^1 E \{ \tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho') \} \tilde{x}(t_i, \rho') d\rho' d\rho \end{bmatrix} \quad (4.243)$$

Evaluating the first term, $P(t_i^-) \tilde{x}(t_i)$, we get

$$\int_0^1 E \{ \tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho') \} \tilde{x}(t_i, \rho') d\rho' \\ = \int_0^1 E \left\{ \sum_{j=0}^{N-1} \epsilon_j(t_i^-) \beta_j(\rho) \sum_{k=0}^{N-1} \epsilon_k(t_i^-) \beta_k(\rho') \right\} \sum_{l=0}^{N-1} \alpha_l(t_i) \beta_l(\rho') d\rho' \quad (4.244)$$

Extending the expectation operator to include the deterministic quantities yields

$$\begin{aligned} & \int_0^1 E \left\{ \tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho') \right\} \tilde{x}(t_i, \rho') d\rho' \\ &= \int_0^1 E \left\{ \sum_{j=0}^{N-1} \epsilon_j(t_i^-) \beta_j(\rho) \sum_{k=0}^{N-1} \epsilon_k(t_i^-) \beta_k(\rho') \sum_{l=0}^{N-1} \alpha_l(t_i) \beta_l(\rho') \right\} d\rho' \end{aligned} \quad (4.245)$$

$$= \int_0^1 E \left\{ \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \epsilon_j(t_i^-) \beta_j(\rho) \epsilon_k(t_i^-) \beta_k(\rho') \alpha_l(t_i) \beta_l(\rho') \right\} d\rho' \quad (4.246)$$

and then we move the expectation operator inside the summations

$$\begin{aligned} & \int_0^1 E \left\{ \tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho') \right\} \tilde{x}(t_i, \rho') d\rho' \\ &= \int_0^1 \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} E \left\{ \epsilon_j(t_i^-) \beta_j(\rho) \epsilon_k(t_i^-) \beta_k(\rho') \alpha_l(t_i) \beta_l(\rho') \right\} d\rho' \end{aligned} \quad (4.247)$$

$$= \int_0^1 \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} E \left\{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \right\} \beta_j(\rho) \beta_k(\rho') \alpha_l(t_i) \beta_l(\rho') d\rho' \quad (4.248)$$

$$= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} E \left\{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \right\} \beta_j(\rho) \alpha_l(t_i) \int_0^1 \beta_k(\rho') \beta_l(\rho') d\rho' \quad (4.249)$$

where in the last line we move all the terms independent of ρ' out in front of the integral. Note that the integral $\int_0^1 \beta_k(\rho') \beta_l(\rho') d\rho'$ is simply the Kronecker delta, thus

$$\begin{aligned} & \int_0^1 E \left\{ \tilde{\mathbf{e}}(t_i^-, \rho) \tilde{\mathbf{e}}(t_i^-, \rho') \right\} \tilde{x}(t_i, \rho') d\rho' \\ &= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} E \left\{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \right\} \beta_j(\rho) \alpha_l(t_i) \delta_{kl} \end{aligned} \quad (4.250)$$

$$= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \alpha_k(t_i) E \left\{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \right\} \beta_j(\rho) \quad (4.251)$$

$$= \boldsymbol{\alpha}^T(t_i) E \left\{ \boldsymbol{\epsilon}(t_i^-) \boldsymbol{\epsilon}^T(t_i^-) \right\} \boldsymbol{\beta}(\rho) \quad (4.252)$$

$$= \boldsymbol{\alpha}^T(t_i) \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\beta}(\rho) \quad (4.253)$$

and therefore the m th element of the vector on the right-hand side of Equation (4.241) is

$$\begin{aligned} & \int_{(m-1)/M}^{m/M} \boldsymbol{\alpha}^T(t_i) E \{ \boldsymbol{\epsilon}(t_i^-) \boldsymbol{\epsilon}^T(t_i^-) \} \boldsymbol{\beta}(\rho) d\rho \\ &= \int_{(m-1)/M}^{m/M} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \alpha_k(t_i) E \{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \} \beta_j(\rho) d\rho \end{aligned} \quad (4.254)$$

$$= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \alpha_k(t_i) E \{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \} \int_{(m-1)/M}^{m/M} \beta_j(\rho) d\rho \quad (4.255)$$

$$= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \alpha_k(t_i) E \{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \} \mu_{m,j} \quad (4.256)$$

where the integrated basis, $\mu_{m,j} = \int_{(m-1)/M}^{m/M} \beta_j(\rho) d\rho$, was defined in Equation (4.185).

Rearranging gives

$$\begin{aligned} & \int_{(m-1)/M}^{m/M} \boldsymbol{\alpha}^T(t_i) E \{ \boldsymbol{\epsilon}(t_i^-) \boldsymbol{\epsilon}^T(t_i^-) \} \boldsymbol{\beta}(\rho) d\rho \\ &= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \mu_{m,j} E \{ \epsilon_j(t_i^-) \epsilon_k(t_i^-) \} \alpha_k(t_i) \end{aligned} \quad (4.257)$$

$$= \boldsymbol{\mu}_m^T E \{ \boldsymbol{\epsilon}(t_i^-) \boldsymbol{\epsilon}^T(t_i^-) \} \boldsymbol{\alpha}(t_i) \quad (4.258)$$

$$= \boldsymbol{\mu}_m^T \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\alpha}(t_i) \quad (4.259)$$

Thus,

$$\mathbf{H} P(t_i^-) \tilde{x}(t_i) = \begin{bmatrix} \boldsymbol{\mu}_1^T \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\alpha}(t_i) \\ \vdots \\ \boldsymbol{\mu}_M^T \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\alpha}(t_i) \end{bmatrix} \quad (4.260)$$

$$= \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \vdots \\ \boldsymbol{\mu}_M^T \end{bmatrix} \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\alpha}(t_i) \quad (4.261)$$

$$= \tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\alpha}(t_i) \quad (4.262)$$

It then follows that the first term of Equation (4.241) becomes

$$\begin{aligned}
P(t_i^+) \tilde{x}(t_i) &= \boldsymbol{\alpha}^T(t_i) \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\beta}(\rho) \\
&\quad - \boldsymbol{\beta}^T(\rho) \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\alpha}(t_i) \quad (4.263)
\end{aligned}$$

$$\begin{aligned}
&= \boldsymbol{\alpha}^T(t_i) \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\beta}(\rho) \\
&\quad - \boldsymbol{\alpha}^T(t_i) \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \boldsymbol{\beta}(\rho) \quad (4.264)
\end{aligned}$$

$$= \boldsymbol{\alpha}^T(t_i) \left\{ \tilde{\mathbf{P}}(t_i^-) - \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \right\} \boldsymbol{\beta}(\rho) \quad (4.265)$$

Since Equation (4.265) applies to every state, by operator equality, the error covariance matrix following the measurement update is

$$\boxed{\tilde{\mathbf{P}}(t_i^+) = \tilde{\mathbf{P}}(t_i^-) - \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-)} \quad (4.266)$$

where the propagated error covariance matrix, $\tilde{\mathbf{P}}(t_i^-)$, is defined in Equation (4.162) and computed in Equation (4.181), the measurement distributor matrix, $\tilde{\mathbf{H}}$, is described in Equation (4.192), and the measurement noise covariance, $\mathbf{R}(t_i)$, is assumed known.

4.3.16 Summary. The previous section began by describing the deterministic heat equation. Then we proceeded to develop the infinite-dimensional, continuous-time dynamics model of the heat equation and a sampled-data measurement model, which led to the creation of the equivalent infinite-dimensional, discrete-time dynamics model. At the end of the section we described how we could derive an essentially-equivalent finite-dimensional discrete-time model from the equivalent infinite-dimensional, discrete-time model. In this section we have systematically determined the finite-dimensional matrix approximations and representations of the multitude of components that comprise the ISKF in such a fashion that we have

synthesized a sampled-data Kalman filter to estimate optimally a finite number of the coefficients associated with a Fourier series expansion of the true state temperature function of the slender cylindrical rod. In preparation for simulating an MMAE to estimate the temperature of the rod adaptively, we summarize the important equations for this sampled-data Kalman filter.

The state coefficients are propagated by

$$\hat{\alpha}(t_{i+1}^-) = \tilde{\Phi}(\Delta t_{i+1}) \hat{\alpha}(t_i^+) + \tilde{\mathbf{B}}_d(t_i) \boldsymbol{\nu}(t_i) \quad (4.131)$$

with error covariance

$$\tilde{\mathbf{P}}(t_{i+1}^-) = \tilde{\Phi}(\Delta t_{i+1}) \tilde{\mathbf{P}}(t_i^+) \tilde{\Phi}(\Delta t_{i+1}) + \tilde{\mathbf{Q}}_d(t_i) \quad (4.181)$$

where the state transition matrix, $\tilde{\Phi}(\Delta t_{i+1})$, for $\Delta t_{i+1} = t_{i+1} - t_i$, is N by N with diagonal elements

$$\left[\tilde{\Phi} \right]_n(t_{i+1} - t_i) = e^{-\kappa n^2 \pi^2 (t_{i+1} - t_i)}, n = 0, 1, \dots, N-1 \quad (4.112)$$

the equivalent discrete-time input distributor is an N by N diagonal matrix with elements

$$\left[\tilde{\mathbf{B}}_d \right]_n(t_i) = \begin{cases} B[t_{i+1} - t_i], & n = 0 \\ \frac{B \left[1 - e^{-\kappa n^2 \pi^2 (t_{i+1} - t_i)} \right]}{\kappa n^2 \pi^2}, & n = 1, 2, \dots, N-1 \end{cases} \quad (4.124)$$

and the equivalent discrete-time system dynamics noise covariance is an N by N diagonal matrix with elements

$$\left[\tilde{\mathbf{Q}}_d \right]_n(t_i) = \begin{cases} Q[t_{i+1} - t_i], & n = 0 \\ \frac{Q \left[1 - e^{-2\kappa n^2 \pi^2 (t_{i+1} - t_i)} \right]}{2\kappa n^2 \pi^2}, & n = 1, 2, \dots, N-1 \end{cases} \quad (4.146)$$

Since $\tilde{\Phi}$ and $\tilde{\mathbf{B}}_d$ are diagonal matrices, the approximate state function coefficient estimates, $\hat{\alpha}_0, \dots, \hat{\alpha}_{N-1}$, are independently propagated; using Equation (4.131) we can write the propagation of the n th coefficient as

$$\hat{\alpha}_n(t_i^-) = \left[\tilde{\Phi} \right]_n (t_i - t_{i-1}) \hat{\alpha}_n(t_{i-1}^+) + \left[\tilde{\mathbf{B}}_d \right]_n (t_i) \nu_n(t_i) \quad (4.267)$$

The measurement update of the approximate state function coefficient estimates is performed by

$$\hat{\boldsymbol{\alpha}}(t_i^+) = \hat{\boldsymbol{\alpha}}(t_i^-) - \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{r}}(t_i) \quad (4.227)$$

where the approximate residual is given by

$$\tilde{\mathbf{r}}(t_i) = \mathbf{z}_i - \tilde{\mathbf{H}} \hat{\boldsymbol{\alpha}}(t_i^-) \quad (4.222)$$

with error covariance given as

$$\tilde{\mathbf{P}}(t_i^+) = \tilde{\mathbf{P}}(t_i^-) - \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T \left[\tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \tilde{\mathbf{H}}^T + \mathbf{R}(t_i) \right]^{-1} \tilde{\mathbf{H}} \tilde{\mathbf{P}}(t_i^-) \quad (4.266)$$

where

$$\tilde{\mathbf{H}} = \begin{bmatrix} \mu_{1,0} & \mu_{1,1} & \cdots & \mu_{1,N-1} \\ \vdots & \vdots & & \vdots \\ \mu_{M,0} & \mu_{M,1} & \cdots & \mu_{M,N-1} \end{bmatrix} \in \mathbb{R}^{M \times N} \quad (4.192)$$

and the elements of $\tilde{\mathbf{H}}$ are

$$\mu_{m,n} = \begin{cases} \frac{1}{M}, & n = 0 \\ \frac{\sqrt{2}}{n\pi} \sin(n\pi\rho) \Big|_{(m-1)/M}^{m/M}, & n = 1, 2, \dots, N-1 \end{cases} \quad (4.195)$$

4.4 Multiple Model Adaptive Estimation

Now that we have determined the finite-dimensional matrix representations needed to implement the ISKF using a digital computer algorithm, next we shall modify the MMAE framework, that was reviewed in Chapter II and extended in Chapter III, in order to implement the MMAE in a simulation of this example in Chapter V.

The formation of the elemental filters was discussed in Section 2.3.3.3. Recall that each elemental filter in the bank (shown in Figure 2.1 on page 2-8) is based upon a different hypothesis for a parameter value used to model the real world system, i.e., the k th elemental filter is constructed assuming that $\mathbf{a}(t_i) = \mathbf{a}_k$, where \mathbf{a}_k is a member of discrete set \mathbb{A} which is a subset of \mathbb{R}^J , the J -dimensional real vector space; J specifies the number of parameters that are adaptively estimated.

In this research, we assume that the prior probability is $p_k(t_0) = 1/K$ for $k = 1, \dots, K$ and that the hypothesis conditional probability for all K hypotheses [125, 107, 11, 108, 130, 132] is computed recursively by

$$p_k(t_i) = \frac{f_{\mathbf{z}(t_i)|\mathbf{a}(t_i), \mathbf{z}(t_{i-1})}(\mathbf{z}_i|\mathbf{a}_k, \mathbf{Z}_{i-1}) p_k(t_{i-1})}{\sum_{j=1}^K f_{\mathbf{z}(t_i)|\mathbf{a}(t_i), \mathbf{z}(t_{i-1})}(\mathbf{z}_i|\mathbf{a}_j, \mathbf{Z}_{i-1}) p_j(t_{i-1})} \quad (4.268)$$

where the conditional PDF,

$$f_{\mathbf{z}(t_i)|\mathbf{a}(t_i), \mathbf{z}(t_{i-1})}(\mathbf{z}_i|\mathbf{a}_k, \mathbf{Z}_{i-1}) = \beta_k(t_i) \exp \left\{ -\frac{1}{2} L_k(t_i) \right\} \quad (4.269)$$

is a zero-mean Gaussian with covariance $\tilde{\mathbf{A}}_k(t_i)$. The PDF scale factor is given by

$$\beta_k(t_i) = \frac{1}{(2\pi)^{M/2} |\tilde{\mathbf{A}}_k(t_i)|^{1/2}} \quad (4.270)$$

with measurement dimension M , and the likelihood quotient, which is a measure of the “correctness” of the parameter values for this particular model [130], being

$$L_k(t_i) = \tilde{\mathbf{r}}_k^T(t_i) \tilde{\mathbf{A}}_k^{-1}(t_i) \tilde{\mathbf{r}}_k(t_i) \quad (4.271)$$

where $\tilde{\mathbf{r}}_k(t_i)$ and $\tilde{\mathbf{A}}_k(t_i)$ are the approximate measurement residual and approximate measurement residual covariance calculated by the k th elemental filter. Note that the primary purpose of $\beta_k(t_i)$ is to scale the PDF so that it integrates to one and that the important information to be retrieved from this PDF is contained in $L_k(t_i)$.

The blended state function coefficient estimate is

$$\hat{\boldsymbol{\alpha}}_{\text{MMAE}}(t_i^+) = \sum_{k=1}^K \hat{\boldsymbol{\alpha}}_k(t_i^+) p_k(t_i) \quad (4.272)$$

where $\hat{\boldsymbol{\alpha}}_k(t_i^+)$ is the state function coefficient estimate generated by the k th elemental filter based on the assumption that the parameter vector $\mathbf{a}(t_i) = \mathbf{a}_k$. The conditional covariance of the state function coefficients is

$$\begin{aligned} & \tilde{\mathbf{P}}_{\text{MMAE}}(t_i^+) \\ &= \sum_{k=1}^K \left\{ \tilde{\mathbf{P}}_k(t_i^+) + [\hat{\boldsymbol{\alpha}}_k(t_i^+) - \hat{\boldsymbol{\alpha}}_{\text{MMAE}}(t_i^+)] [\hat{\boldsymbol{\alpha}}_k(t_i^+) - \hat{\boldsymbol{\alpha}}_{\text{MMAE}}(t_i^+)]^T \right\} p_k(t_i) \end{aligned} \quad (4.273)$$

where $\tilde{\mathbf{P}}_k(t_i^+)$ is the state coefficient error covariance computed by the k th elemental filter. Since we are really interested in the temperature state function (versus a vector of coefficients), use Equation (4.34) to recreate the state:

$$\hat{x}(t_i, \rho) = \hat{\boldsymbol{\alpha}}^T(t_i) \boldsymbol{\beta}(\rho) \quad (4.274)$$

The parameter estimate is given by

$$\hat{\mathbf{a}}_{\text{MMAE}}(t_i^+) = \sum_{k=1}^K \mathbf{a}_k p_k(t_i) \quad (4.275)$$

with conditional covariance of $\mathbf{a}(t_i)$ [129]:

$$\mathbf{P}_{\mathbf{a}, \text{MMAE}}(t_i^+) = \sum_{k=1}^K [\mathbf{a}_k - \hat{\mathbf{a}}_{\text{MMAE}}(t_i^+)] [\mathbf{a}_k - \hat{\mathbf{a}}_{\text{MMAE}}(t_i^+)]^T p_k(t_i) \quad (4.276)$$

4.5 *Summary*

The purpose of this chapter was to demonstrate, using a practical straightforward example, the filtering theory developed in Chapter III. The temperature along a slender cylindrical rod was observed using a sampled-data measurement model and an essentially-equivalent finite-dimensional discrete-time model (derived from an infinite-dimensional, continuous-time model for the flow of heat — a scalar stochastic heat equation) was employed to characterize the system dynamics for this practical problem. In the bulk of this chapter, we systematically found the finite-dimensional matrix approximations and representations for the multitude of components that comprise the ISKF in such a fashion that we have synthesized a sampled-data Kalman filter to estimate optimally a finite number of the coefficients associated with a Fourier series expansion of the true state temperature function of the slender cylindrical rod. Finally, we crafted a fixed-bank MMAE composed of these finite-dimensional filters to estimate the temperature along the rod. In the sequel to this chapter, we shall use a Monte Carlo simulation to investigate this example further.

V. *Simulation and Results*

5.1 *Introduction*

This chapter discusses the results of six computer-based Monte Carlo simulations¹ in detail. The first five simulations feature state and/or parameter estimation in the presence of an uncertain noise environment. Specifically, the first two simulations involve estimating the covariance of the dynamics driving noise; this can be a rather difficult task if the measurements are of relatively poor quality as compared to the quality of the dynamics model. The quality (or precision) of the dynamics model is expressed in terms of the noise covariance, \mathbf{Q}_d , while the quality of the observation model is related by the measurement-corruption noise covariance, \mathbf{R} . In the first two simulations, we indirectly estimate the dynamics noise covariance using noise-corrupted measurements. In the next three simulations the measurement-corruption noise covariance is found. The fourth simulation demonstrates the MMAE's capability to adjust to a linearly changing measurement-corruption noise covariance — both increasing and decreasing over time. In the fifth simulation, the elemental filters in the MMAE demonstrate their ability to change their status from poorly modeling the real-world environment to correctly modeling it in just a few short propagate/update cycles as the truth model measurement noise covariance abruptly changes (twice) during the course of the simulation. The capability to identify an unknown system parameter, often called system identification [9], is addressed in the final simulation discussed in this chapter. The MMAE demonstrates its ability to identify a system parameter (in this case, a material property of the slender cylin-

¹Maybeck's [134] MATLAB-based MMAE software was modified for use in this dissertation to create an approximate infinite-dimensional MMAE (AIMMAE). A technical report containing all of the code necessary to duplicate the results reported herein is available upon request from either Dr. Peter Maybeck at Peter.Maybeck@AFIT.edu or Dr. Scott Sallberg at Sallberg@IEEE.org.

drical rod), and from that we can ascertain the most likely rod material from a short list of fairly distinct materials.

All six simulations decomposed the temperature function into (at least) thirty “states” for both the truth model and elemental filter models. Recall that these states are actually the coefficients, $\{\alpha_0(t_i), \alpha_1(t_i), \dots, \alpha_{29}(t_i)\}$, that correspond to the basis elements, $\{1, \sqrt{2} \cos(\pi\rho), \dots, \sqrt{2} \cos(29\pi\rho)\}$, used to decompose the (possibly) infinite-dimensional state function; see the discussion in Section 4.2.8 on page 4-12. With thirty basis elements, the mean-squared error (MSE) between the actual temperature along the rod (for a ramp) and the Gibbs effect resulting from the finite number of basis elements used to represent the signal are both small relative to the other choices investigated; ten, fifteen, twenty, thirty, forty, fifty, sixty, seventy, eighty, and one hundred basis elements were initially used in this preliminary study. Along another line of reasoning employing the state transition matrix, given in Equation (4.112) on page 4-28, we may reasonably expect that only a small number of state coefficients are needed to model the state function adequately. For instance, given that the diagonal state transition matrix has elements: $\left[\tilde{\Phi}\right]_n(t_{i+1} - t_i) = e^{-\kappa n^2 \pi^2 (t_{i+1} - t_i)}$ and choosing $\kappa = 1 \text{ m}^2/\text{sec}$ and $t_{i+1} - t_i = 0.01 \text{ sec}$, we get: $\left[\tilde{\Phi}\right]_{10}(t_{i+1} - t_i) = 5.2 \times 10^{-5}$, $\left[\tilde{\Phi}\right]_{15}(t_{i+1} - t_i) = 2.3 \times 10^{-10}$, $\left[\tilde{\Phi}\right]_{20}(t_{i+1} - t_i) = 7.2 \times 10^{-18}$, and $\left[\tilde{\Phi}\right]_{30}(t_{i+1} - t_i) = 2.6 \times 10^{-39}$. Thus, $N = 15$ might be all that we need, since coefficients past the fifteenth are reduced by a factor of a billion or more during each propagation cycle. Depending on the computer resources available, computational loading may also limit the number of states (coefficients) employed in the model. Finally, since the real truth model would be an infinite-dimensional model, we have, in a sense, a reduced-order model.

All of the simulations feature a slender cylindrical rod partitioned into five segments; a separate measurement is taken over each segment. The five measurements are recorded together at each time instant in the measurement vector. Preliminary studies indicated that five segments was a nice balance between too few and too

Parameter	Symbol	Value
Thermal diffusivity constant	κ	1 m ² /sec
Number of measurements (sensors)	M	5
Number of states (basis elements)	N	30
System dynamics noise strength	Q	5 (°C) ² /sec
Measurement-corruption noise covariance	R	5 (°C) ²
Initial state (temperature along rod)	\hat{x}_0	20 °C
Initial state covariance	P_0	25 (°C) ²

Table 5.1 Truth Parameters

many segments. Fewer segments would have resulted in a coarser spatial discretization, which would have made it more difficult to detect the onset of excitation (heat) in simulation six and thus degrade system identification capabilities. However, we did not want to describe the temperature perfectly at every point along the rod with too many segments either.

Fifty Monte Carlo runs were used in each simulation to generate the hundreds of plots seen on the dozens of figures on the following pages. Once again, we note that all of these plots were created from a set of fifty Monte Carlo runs that utilized a clock-based seeding of the random number generator. Several of the pertinent truth model parameters are stated in Table 5.1.

Notes pertaining to Table 5.1:

1. We have chosen to work directly with the dynamics noise strength, Q , since it is part of the original continuous-time model description, see Section 4.2.1, versus the equivalent discrete-time diffusion, Q_d , which is calculated by the software as needed.
2. Maybeck [129] discusses at length the tuning of elemental filters by adjusting the Q/R ratio. When these noise parameters are matrices, then the talk shifts to a ratio of their largest eigenvalues — the same trends still apply although it is somewhat more complicated if the matrices are not diagonal matrices. With

that in mind, we note that the truth measurement-corruption noise covariance is a diagonal matrix $\mathbf{R} = 5\mathbf{I}$, where \mathbf{I} is an M -by- M identity matrix. The eigenvalues of \mathbf{R} are all equal to five. Hence, when we say R , we are referring to the largest (and only) eigenvalue of \mathbf{R} ; however, we will call both of them the *measurement-corruption noise covariance* and the context will dictate whether we are referring to the matrix or the largest eigenvalue.

3. The eigenvalues of the initial state covariance \mathbf{P}_0 are all equal to twenty-five. Hence, when we say P_0 , we are referring to the largest (and only) eigenvalue of \mathbf{P}_0 ; however, we will call both of them the *initial state covariance* and the context will dictate whether we are referring to the matrix or the largest eigenvalue.
4. In the fourth simulation, the truth R is varied linearly, both increasing and decreasing.
5. In the fifth simulation, we change the truth measurement-corruption noise covariance twice in an abrupt fashion.
6. For the first five simulations, the thermal diffusivity constant is set to $\kappa = 1$, which for comparison purposes places it halfway between aluminum ($\kappa = 0.86$) and copper ($\kappa = 1.14$). In the sixth simulation, we set the thermal diffusivity constant to $\kappa = 0.86$ to perform system identification.

For each of the simulations, a short description sets up the goals and explains a few pertinent facts about the truth model, some of which are tabulated if they differ from Table 5.1. Additionally, we include a graphical description of the parameter set to help assess how the elemental filters in the filter bank differ; see, for example, Figure 5.12 on page 5-29. The simulation results feature a large collection of figures. For the last five simulations, 2 through 6, a figure for initial results for each simulation, such as Figure 5.13 on page 5-30 for an investigation into initial state covariance settings, displays the overall probability flow among the elemental filters. This plot is intended to emphasize the MMAE's indication of the filter

(a) Temp at ‘left’ end of rod ($\rho = 0$)	(b) Error temp at ‘left’ end of rod ($\rho = 0$)
(c) Temp at ‘center’ of rod ($\rho = 0.5$)	(d) Error temp at ‘center’ of rod ($\rho = 0.5$)
(e) Temp at ‘right’ end of rod ($\rho = 1$)	(f) Error temp at ‘right’ end of rod ($\rho = 1$)
(g) Likelihood quotient: $\mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k^T(t_i)$	(h) Hypothesis conditional probability: $p_k(t_i)$

Table 5.2 Arrangement for Plots (a) through (h) for the k th Elemental Filter

based on the best hypothesis², and thus the elemental filter(s) *most responsible* for the overall MMAE performance. In this dissertation, the *best* hypothesis is defined as the hypothesis that gives rise to the elemental filter with the largest hypothesis conditional probability; hence, the *best* elemental filter is the filter which receives the highest hypothesis conditional probability. Following the introduction to the simulation and a brief discussion of expected filter performance, there are two pages of plots for each of the (three or five) elemental filters³ in the MMAE filter bank, followed by a two-page set of plots for the blended filter. The first set of plots for an elemental filter contains a full accounting of the filter’s progression through time at three strategic points along the rod in plots (a) through (f); a plot of the likelihood quotient appears in plot (g), and plot (h) contains the hypothesis conditional probability — see Table 5.2 for the placement of these plots in the figure. Plot (h) contains a comparison of the mean plus and minus one standard deviation of the hypothesis conditional probability for each of the elemental filters in the bank for each simulation; this plot enables a quick analysis of the probability flow among the elemental filters.

²Equation (4.268), on page 4-55, is used to calculate these probabilities.

³For the first simulation, we do not give an individual accounting of the elemental filters since the focus of the first simulation was to discover/identify trends regarding the filter bank composition. With that in mind, the figures for simulation one are composite summaries of the hypothesis conditional probability histories for all five elemental filters in the filter bank.

For plots (a), (c), and (e), the solid line represents the mean elemental filter estimate (a mean taken over the 50 Monte Carlo runs) while the dashed line represents the truth state. In plots (b), (d), and (f), the solid line represents the mean elemental filter error, the dash-dot lines are at plus and minus one truth standard deviation from the elemental filter mean error (solid line), and dashed lines are zero (the filter-assumed mean error) plus and minus filter-computed standard deviation. Plot (g) has a solid line for the mean value of the likelihood quotient: $L_k(t_i) = \mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)$, a measure of the “correctness” of the parameter values for the k th elemental filter [130, 132]; the dash-dot lines are at plus and minus one truth standard deviation. In plot (h), the solid line traces out the mean hypothesis conditional probability, $p_k(t_i)$, the probability that the assumed parameter value is correct conditioned on the observed measurement history through time t_i [130, 132]; the dash-dot lines are at plus and minus one standard deviation.

In practice, it is useful to inspect plot (h) for each of the elemental filters first; that is why we have included a summary of the (h) plots for the entire filter bank early in the discussion (namely, in the “initial results” figure discussed on page 5-4). The “initial results” summary tells us which filter is most responsible for the overall MMAE performance and it directs our attention to the performance of the elemental filter with the best hypothesis. [For simulation one, only initial results are reported; they are given in figures containing sixteen such “initial results” plots arranged in a four by four array; see, for example Figure 5.1 on page 5-15.] This initial results summary does not replace the (h) plots — we still need plot (h) because it contains more information about the performance; plot (h) tells us how much the mean hypothesis conditional probability varies over the 50 Monte Carlo runs by stating the mean plus and minus one standard deviation. After considering plot (h), we usually look at the likelihood quotient in plot (g). It is highly likely that the best filter model (the one with the largest mean hypothesis conditional probability) will also display a sequence of likelihood quotients equal to the number of sensor

segments, while the filters with the lowest mean hypothesis conditional probability will often have a sequence of likelihood quotients much larger than the number of sensor segments. Next, the state estimate in plots (a), (c), and (e) are assessed more fully by the estimation error in plots (b), (d), and (f). Among other attributes of the elemental filter, the adequacy of the initial state covariance can be checked by inspecting plots (b), (d), and (f); the initial error should be within the 1σ (i.e., one standard deviation) bounds created by the initial state covariance. If this is not true, then convergence is greatly hampered since the filter has been told that its initial condition errors are much smaller than they really are. Finally, we repeat this inspection process for the other “less probable” elemental filters.

Plot (g), of for example Figure 5.14, on page 5-36, features the likelihood quotient. For a properly matched elemental filter, we expect a likelihood quotient of around $M = 5$. This is not a guarantee that the filter is properly matched because the filter-computed residual covariance could be masking poor residuals; however, a likelihood quotient significantly greater than $M = 5$ is a strong indication that the filter is mismodeled. The best indication of a filter based on the correct model for the situation is given by the hypothesis conditional probability in plot (h). When $p_k = 1$, the MMAE is indicating that the model *completely* matches reality with probability one; here, the *completeness* is relative to the other elemental filters which are based on relatively inaccurate models as compared to the elemental filter receiving $p_k = 1$. When p_k is small, the filter is either mismatched or poorly tuned.

Improved state estimation is usually the end goal of our adaptive estimation process, i.e., we seek to produce a better state estimate using an MMAE structure, rather than a precise estimate of the uncertain parameter itself. However, in Section 5.7, we use the MMAE to estimate the parameter of interest — also known as system identification. The MMAE state estimator usually outperforms a similar estimator based on a single elemental filter (based on a single assumed value for the uncertain parameter). The second set of plots, (i) through (p), for each elemental filter displays

the temperature along the entire length of the rod at selected instances of time from time zero to the end of the simulation. The state at time zero reflects the initial conditions, while the state at the end of the simulation includes the final measurement update. With the exception of the first time instant, all of the times show results for just after the measurement update, i.e., at time t_i^+ . Additionally, the root mean square (RMS) error of the temperature estimate is displayed on each subplot to help quantify the performance over the entire rod⁴. The solid line represents the filter's mean temperature estimate while the dashed line is the true temperature.

The first set of plots for the blended filter⁵ is similar to plots (a) through (f) for an elemental filter, including the dashed lines in (b), (d), and (f) that reflect the blended filter-indicated zero plus and minus one sigma values. Plot (g) shows the RMS error, as a function of time, for the MMAE blended state estimate. The second set of plots, (h) through (o), for the blended filter, are nearly the same as the second set for the elemental filters. While the elemental filter time line begins with the initial conditions, the blended filter time line begins after the MMAE has produced its first estimate of the temperature, i.e., after the first measurement update — the blended filter represents the state estimate computed using Equation (4.272) on page 4-56 and its conditional covariance in Equation (4.273). Thus, a careful inspection of plots (i) for each of the elemental filters and plot (h) for the blended filter will show that they do in fact, not reflect the exact same instant of time at $t_i = 0$ sec. However, the remaining temperature snapshots *are* for the same time instances.

It should not be surprising to the reader to find on the following pages, better adaptation to uncertainty in the measurement-corruption noise covariance versus system-dynamics noise strength. Consider the fact that, first of all, models for

⁴The RMS error on these plots is an “instantaneous” value and should not be treated as an absolute indicator that can be compared exactly from plot to plot, and especially not between different experiments, nor between cases within a particular simulation, since the results are based on different sets of Monte Carlo runs. The trends, however, are certainly valid.

⁵The blended filter is not really a filter, but a blending of the data/results from the elemental filters.

dynamic system processes are often less precise or less well understood as compared to our measurement models for a particular problem or application. Additionally, while we can calibrate and inspect our measurement apparatus, we often cannot do likewise for the system itself. Furthermore, knowledge of the dynamics is only available to us through the measurement process; a process which adds yet another layer of uncertainty to our state estimation. More specifically, the uncertainties in the measurement-corruption noise \mathbf{v} , as in its covariance \mathbf{R} , directly impact the measurement residuals since $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$. On the other hand, the effects of dynamics driving noise parameters, such as Q , first impact the system dynamics and (after the inherent delays and other effects of the system itself) are then reflected in the state values as expressed by $\mathbf{H}\mathbf{x}$ in the measurement equation $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$. Thus, the effects of uncertain dynamics model parameters are not as directly viewable as those in the measurement model.

Before delving into the results of the simulations, we shall investigate the expected behavior for the likelihood quotient: $L_k(t_i) = \mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)$, where the filter-computed residual covariance, $\mathbf{A}_k(t_i) = \mathbf{H}_k(t_i) \mathbf{P}_k(t_i^-) \mathbf{H}_k^T(t_i) + \mathbf{R}_k(t_i)$, depends on the elemental filter design model. In the third, fourth, and fifth simulations, we have five elemental filters, all with different measurement-corruption noise covariances. Thus, we have five different \mathbf{A}_k matrices. We can use this knowledge to help us predict what the likelihood quotient will be for all of the elemental filters before we run any simulation. Recall that when an elemental filter matches the true scenario, the history of residuals form a zero-mean white Gaussian noise sequence with known covariance — the filter-computed residual covariance \mathbf{A} — and the expected likelihood quotient is equal to the number of measurements, M , and in practice, this is generally the case. When the filter is mismatched, however, L_k will usually differ considerably (and often becomes several orders of magnitude larger than the number of measurements).

From Maybeck [129] we know that the zero-mean residual sequence for a Kalman filter based upon the true parameter value has covariance

$$E\{\mathbf{r}(t_i) \mathbf{r}^T(t_i) | \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\} = \mathbf{A}_{\text{true}}(t_i) \quad (5.1)$$

which we note is independent of the measurement history \mathbf{Z}_{i-1} . Now we find the expected value of the random likelihood quotient, $\mathbf{L}_k(t_i)$, at time t_i for the k th elemental filter

$$E\{\mathbf{L}_k(t_i)\} = E\{\mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)\} = \text{tr}\{E[\mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)]\} \quad (5.2)$$

where $\mathbf{r}_k(t_i)$ is the random measurement residual; additionally, we note that the trace of a scalar function is equal to that scalar function. Moving the trace inside the expectation yields

$$E\{\mathbf{L}_k(t_i)\} = E\{\text{tr}[\mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)]\} \quad (5.3)$$

and this formulation allows the terms to commute, thus we obtain

$$E\{\mathbf{L}_k(t_i)\} = E\{\text{tr}[\mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i) \mathbf{r}_k^T(t_i)]\} \quad (5.4)$$

Next we move the expectation back in and get

$$E\{\mathbf{L}_k(t_i)\} = \text{tr}\{\mathbf{A}_k^{-1}(t_i) E[\mathbf{r}_k(t_i) \mathbf{r}_k^T(t_i)]\} \quad (5.5)$$

Using Equation (5.1)

$$E\{\mathbf{L}_k(t_i)\} = \text{tr}\{\mathbf{A}_k^{-1}(t_i) \mathbf{A}_{\text{true}}(t_i)\} \quad (5.6)$$

If for all time t_i , $\mathbf{A}_k(t_i) = \mathbf{A}_{\text{true}}(t_i)$, then the right-hand side of Equation (5.6) becomes the trace of the identity matrix which is equal to the dimension of the

matrix, in this case M . Thus, in the simulations considered here, we expect the likelihood quotient to be five, the number of sensor segments, when the filter hypothesis is *correct*. However, most of the filters are created using an incorrect hypothesis and thus we usually have $\mathbf{A}_k(t_i) \neq \mathbf{A}_{\text{true}}(t_i)$, where we recall that $\mathbf{A}(t_i) = \mathbf{H}(t_i) \mathbf{P}(t_i^-) \mathbf{H}^T(t_i) + \mathbf{R}(t_i)$. In steady state operation, which usually occurs after just a few propagate/update cycles, we have $\mathbf{A}_k^{\text{ss}} \approx \mathbf{R}_k$ and $\mathbf{A}_{\text{true}}^{\text{ss}} \approx \mathbf{R}_{\text{true}}$ for the particular example simulated herein. Thus,

$$E\{\mathbf{L}_k(t_i)|_{t_i=t_{\text{ss}}}\} = \text{tr}\{\mathbf{R}_k^{-1}\mathbf{R}_{\text{true}}\} \quad (5.7)$$

$$= \text{tr}\{[R_k\mathbf{I}]^{-1}R_{\text{true}}\mathbf{I}\}, \text{ where } \mathbf{R}_k = R_k\mathbf{I} \text{ and } \mathbf{R}_{\text{true}} = R_{\text{true}}\mathbf{I} \quad (5.8)$$

$$= \frac{R_{\text{true}}}{R_k} \text{tr}\{\mathbf{I}\} \quad (5.9)$$

$$= \frac{R_{\text{true}}}{R_k} M \quad (5.10)$$

where we assume that (1) all M of the eigenvalues of \mathbf{R}_k are equal to R_k , (2) all M of the eigenvalues of \mathbf{R}_{true} are equal to R_{true} , and (3) M is the dimension of the square covariance matrices \mathbf{R}_k and \mathbf{R}_{true} and the identity matrix \mathbf{I} . The first equality in the above development is due to the steady state assumption; the fourth equal sign is by the definition of the trace of a matrix — it is the sum of the diagonal elements. Therefore, we can use Equation (5.10) to predict the steady state likelihood quotient for an elemental filter created using a model that differs from the true value of R ; this will be very valuable for our analysis in Simulations 3, 4, and 5.

In the first five simulations, we initially investigate state estimation in the presence of an unknown noise environment and then (in the sixth simulation) we perform system identification. We find that, even when we “poorly” identify the system-dynamics noise strength, the state estimation task fairs well. We treat filter bank composition and filter initialization issues for improving adaptation to unknown noise statistics.

5.2 Simulation 1

When a Kalman filter is designed using mismatched noise statistics, the filter is no longer optimal [78, 143, 129, 169]. Thus we are motivated in this first simulation to demonstrate how a bank of filters can be used to estimate the value of an unknown parameter — the system dynamics driving noise strength Q . While we are specifically calling attention to this one parameter, we are really interested in improving our state estimation and identifying trends useful for filter bank composition. Initial work on this simulation showed that it was difficult to create a bank of elemental filters that both spanned the entire range of expected values and appeared distinct from one another. Thus, before we can concentrate on state estimation, we must first create a good bank of filters.

As a precursor to using the MMAE to estimate the dynamics noise strength, Q , we conduct a study to find the best discrete set of values to represent the continuum of possible values for Q . The objective of this experiment is not to estimate the parameter, but to illustrate the parameter space discretization as indicated by the hypothesis conditional probability time histories. For optimal state and/or parameter estimation performance, the elemental filters in the filter bank must be distinguishable from one another. In addition to being distinct, the elemental filters must be based on a set of parameter values that covers the entire range of expected values for the parameter of interest.

Tables 5.3 and 5.4 contain key truth and filter parameter values⁶ for the 16-case runs displayed in Figures 5.1 to 5.11. The elemental filter quintet for each case displayed in Figures 5.1 to 5.11 is determined using the data in Table 5.4. As an example, the Q values used to construct the elemental filter quintets for the $Q_{\text{true}} = 10$, $R_{\text{true}} = 0.1$ case are reported in Table 5.5; this corresponds to results

⁶In Table 5.4, as in the rest of the tables in this chapter, important values that are constant across the elemental filters in the bank are only listed once. For example, all three elemental filters in Table 5.4 are designed using the same R , \hat{x}_0 , and P_0 values, whereas the Q is different for each elemental filter.

Figure	N	Q_{true}	R_{true}	R_{filter}
5.1	30	10 $(^{\circ}\text{C})^2/\text{sec}$	0.1 $(^{\circ}\text{C})^2$	0.1 $(^{\circ}\text{C})^2$
5.2	30	100 $(^{\circ}\text{C})^2/\text{sec}$	1 $(^{\circ}\text{C})^2$	1 $(^{\circ}\text{C})^2$
5.3	30	50 $(^{\circ}\text{C})^2/\text{sec}$	1 $(^{\circ}\text{C})^2$	1 $(^{\circ}\text{C})^2$
5.4	30	20 $(^{\circ}\text{C})^2/\text{sec}$	1 $(^{\circ}\text{C})^2$	1 $(^{\circ}\text{C})^2$
5.5	30	10 $(^{\circ}\text{C})^2/\text{sec}$	1 $(^{\circ}\text{C})^2$	1 $(^{\circ}\text{C})^2$
5.6	40	10 $(^{\circ}\text{C})^2/\text{sec}$	1 $(^{\circ}\text{C})^2$	1 $(^{\circ}\text{C})^2$
5.7	50	10 $(^{\circ}\text{C})^2/\text{sec}$	1 $(^{\circ}\text{C})^2$	1 $(^{\circ}\text{C})^2$
5.8	60	10 $(^{\circ}\text{C})^2/\text{sec}$	1 $(^{\circ}\text{C})^2$	1 $(^{\circ}\text{C})^2$
5.9	70	10 $(^{\circ}\text{C})^2/\text{sec}$	1 $(^{\circ}\text{C})^2$	1 $(^{\circ}\text{C})^2$
5.10	30	100 $(^{\circ}\text{C})^2/\text{sec}$	1 $(^{\circ}\text{C})^2$	10 $(^{\circ}\text{C})^2$
5.11	30	100 $(^{\circ}\text{C})^2/\text{sec}$	10 $(^{\circ}\text{C})^2$	1 $(^{\circ}\text{C})^2$

Table 5.3 Simulation 1: Four key parameters for the filter bank composition experiment. N represents the order of the system model.

shown in Figure 5.1. Thus, the far left column of plots contain filter banks centered on the true value of Q , while the second column (from the left) contains filter banks centered on twice the true value of Q . The top row contains filter banks with Q values spaced a decade apart, while the bottom row of filter banks are spaced by two orders of magnitude. Each of the subplots for each case represents the average results for fifty Monte Carlo runs; thus, each of the figures from Figure 5.1 to 5.11 contains the probability flow results for 800 Monte Carlo runs. Finally, the discretization method chosen for this experiment is reminiscent of the simple logarithmic spacing of the elemental filters proposed in Section 2.3.3.

Before we begin our analysis, we shall take a tour of Figure 5.1, the first of eleven such figures. In each column of plots, going down a column corresponds to increasing the coarseness of the discretization; thus we expect that the proper distinguishability of the elemental filters will increase, as demonstrated by an increase in the share of the probability received by the elemental filter designed with the most appropriate value for the dynamics noise strength. The trend observed by looking across a row of plots is not always as straightforward since it is somewhat

Filter	Q_{filter}	R_{filter}	\hat{x}_0	P_0
1	$(c/d^2)Q_{\text{true}}$			
2	$(c/d)Q_{\text{true}}$			
3	cQ_{true}	R_{true}	20	10
4	cdQ_{true}			
5	cd^2Q_{true}			

Table 5.4 Simulation 1: Elemental filter parameters for the filter bank composition experiment. c represents the *centering* of parameter values used as a “basis” for the elemental filters in the bank. d pertains to the *discretization* of the parameter set. Note that R_{filter} for the cases in Figures 5.10 and 5.11 were over- and underestimated by a factor of ten, respectively.

	$c = 1$	$c = 2$	$c = 5$	$c = 10$
$d = 10$	$\frac{1}{10}, 1, 10, 100, 1000$	$\frac{1}{5}, 2, 20, 200, 2000$	$\frac{1}{2}, 5, 50, 500, 5000$	$1, 10, 100, 1000, 10000$
$d = 20$	$\frac{1}{40}, \frac{1}{2}, 10, 200, 4000$	$\frac{1}{20}, 1, 20, 400, 8000$	$\frac{1}{8}, 2\frac{1}{2}, 50, 1000, 20000$	$\frac{1}{4}, 5, 100, 2000, 40000$
$d = 50$	$\frac{1}{250}, \frac{1}{5}, 10, 500, 25000$	$\frac{1}{125}, \frac{2}{5}, 20, 1000, 50000$	$\frac{1}{50}, 1, 50, 2500, 125000$	$\frac{1}{25}, 2, 100, 5000, 250000$
$d = 100$	$\frac{1}{1000}, \frac{1}{10}, 10, 1000, 100000$	$\frac{1}{500}, \frac{1}{5}, 20, 2000, 200000$	$\frac{1}{200}, \frac{1}{2}, 50, 5000, 500000$	$\frac{1}{100}, 1, 100, 10000, 1000000$

Table 5.5 Dynamics noise strengths for each of the elemental filters shown in the 16 plots of Figure 5.1, where $Q_{\text{true}} = 10$ and $R_{\text{true}} = 0.1$. Note that only the first column (and the top right case) have the true value for Q included in the bank of elemental filters. Additionally, the last four elemental filters shown in the top left plot are constructed using the same model as the first four elemental filters of the top right case.

dependent on the discretization level. In the first column of plots, the third (or central) elemental filter in the bank of five filters is designed with artificial knowledge of the true dynamics noise strength; in the second, third, and fourth columns, the central filter is designed for $2Q_{\text{true}}$, $5Q_{\text{true}}$, and $10Q_{\text{true}}$, respectively. Thus, for the case of $(c, d) = (10, 10)$, the second elemental filter assumes (correctly) that the noise strength is Q_{true} , while the central filter is no longer the best filter since it was designed for $10Q_{\text{true}}$. However, in general, the central filter becomes less probable as the multiplicative offset (dictated by c) increases, while the second elemental filter becomes more probable since its underestimate of the noise strength becomes small relative to the overestimate assumed by the central filter.

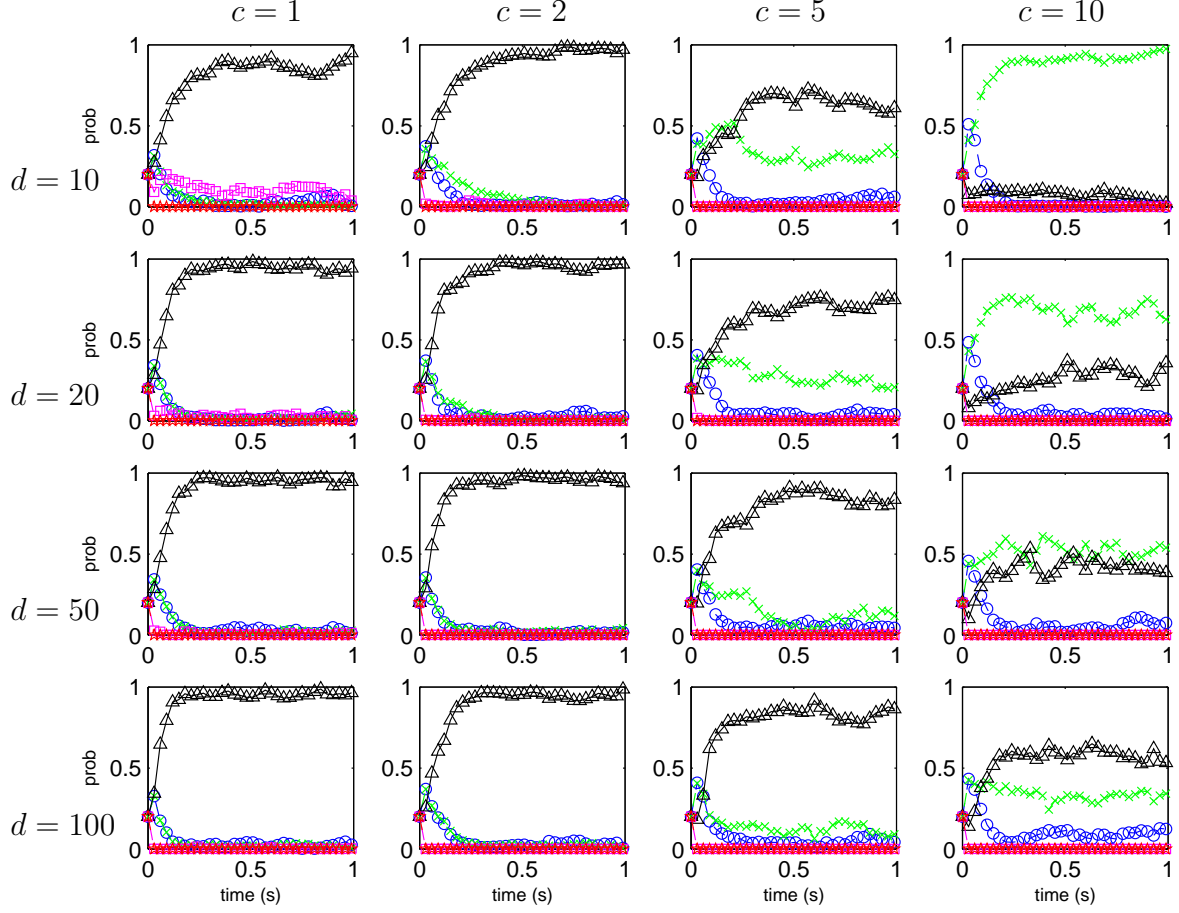


Figure 5.1 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 30$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 0.1$. Filter 1 (\circ): $Q_1 = 10c/d^2$, Filter 2 (\times): $Q_2 = 10c/d$, Filter 3 (Δ): $Q_3 = 10c$, Filter 4 (\square): $Q_4 = 10cd$, Filter 5 (\star): $Q_5 = 10cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

Furthermore, it may be helpful to consider the following simple mnemonic for associating the markers on the hypothesis conditional probability history plots to its respective elemental filter: The *first* filter symbol is a *single* curved line in the shape of a circle \circ . Next, we employ *two* crossed line segments, \times , to represent the *second* elemental filter. The *three* angles in a triangle Δ marks the graph for elemental filter *three*. Elemental filter *four* uses the *four*-sided square \square . Finally, elemental filter *five* employs the *five*-pointed star \star to annotate its probability history.

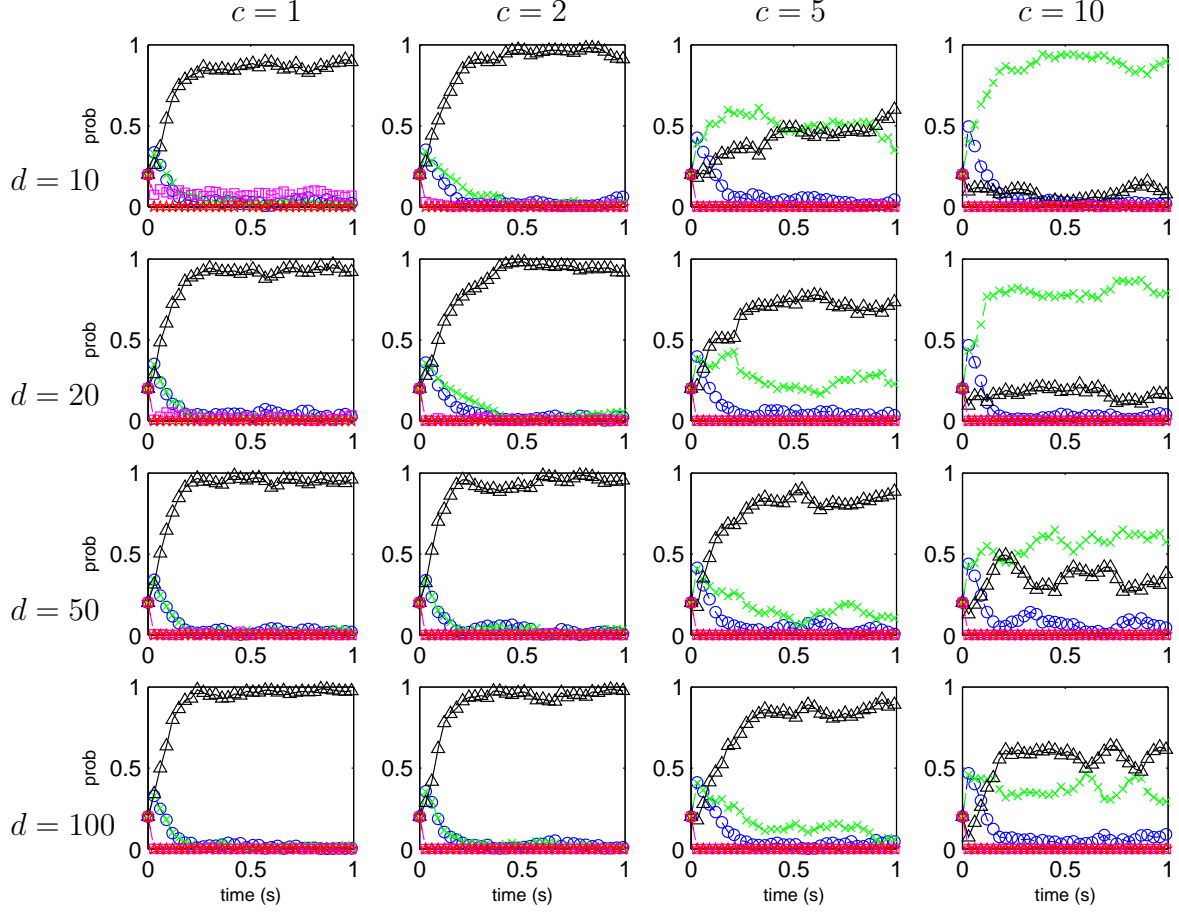


Figure 5.2 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 30$, $Q_{\text{true}} = 100$, $R_{\text{true}} = 1$. Filter 1 (\circ): $Q_1 = 100c/d^2$, Filter 2 (\times): $Q_2 = 100c/d$, Filter 3 (Δ): $Q_3 = 100c$, Filter 4 (\square): $Q_4 = 100cd$, Filter 5 (\star): $Q_5 = 100cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

There are several trends regarding filter bank composition that can be readily seen by inspecting the plots in Figures 5.1 through 5.11. Using the plots in these eleven figures, we shall graphically exhibit three trends which feature increased probability for the elemental filter based on the most correct model: increasing d (i.e., increasing the coarseness of the parameter discretization), increasing the Q/R ratio, where Q is the dynamics noise strength and R is the measurement-corruption noise covariance, and/or increasing the order of the model, N . Our first choice may

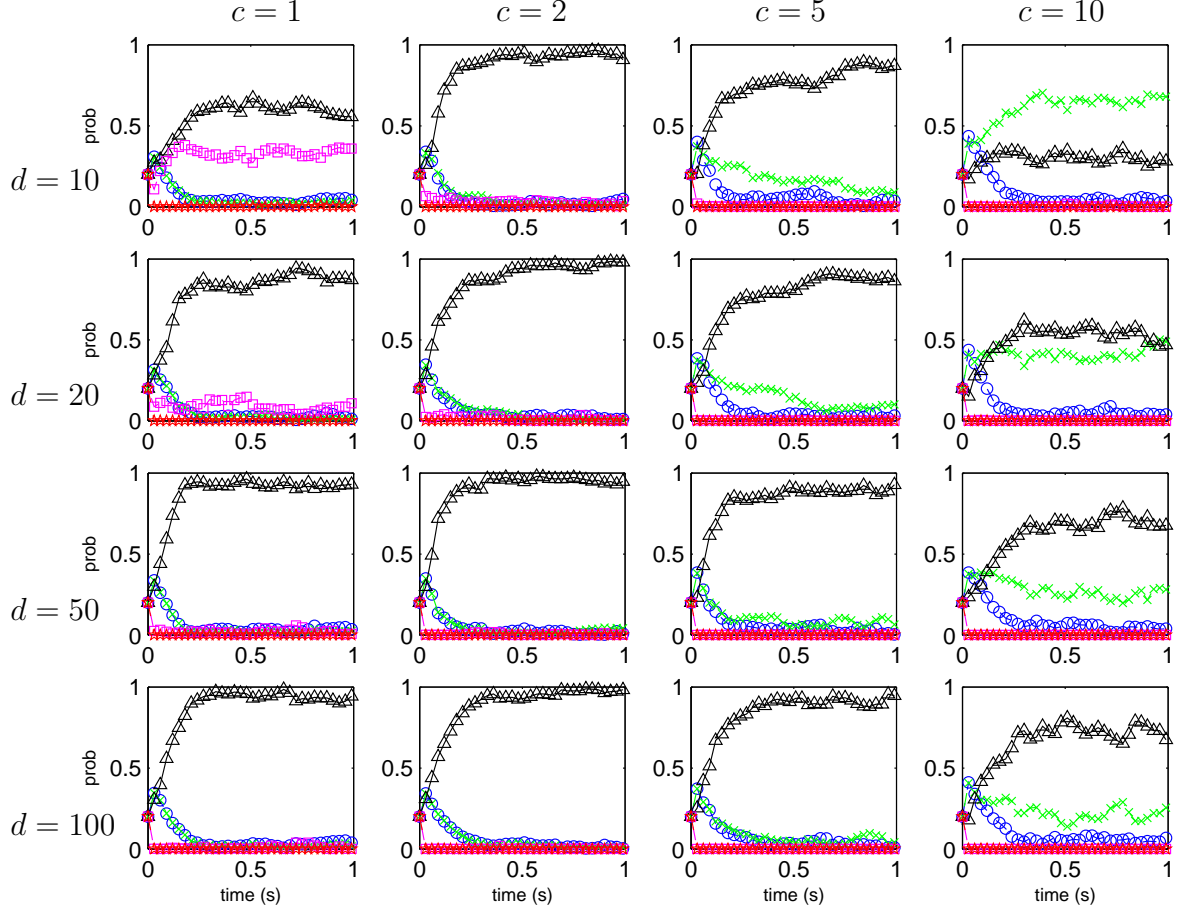


Figure 5.3 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 30$, $Q_{\text{true}} = 50$, $R_{\text{true}} = 1$. Filter 1 (\circ): $Q_1 = 50c/d^2$, Filter 2 (\times): $Q_2 = 50c/d$, Filter 3 (Δ): $Q_3 = 50c$, Filter 4 (\square): $Q_4 = 50cd$, Filter 5 (\star): $Q_5 = 50cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

be to increase the coarseness of the discretization, but, this is a tradeoff that exchanges quality of the state estimate for improved parameter estimation; a moving bank structure, see the discussion in Section 2.6, may help ease these trade-off costs. When increasing N is not affordable computationally, perhaps we can tune the filters and achieve a more favorable probability flow to what we believe may be the most appropriate elemental filter. As we might expect, these results are dependent on correct assumptions for the other model parameters. In the filter-bank results

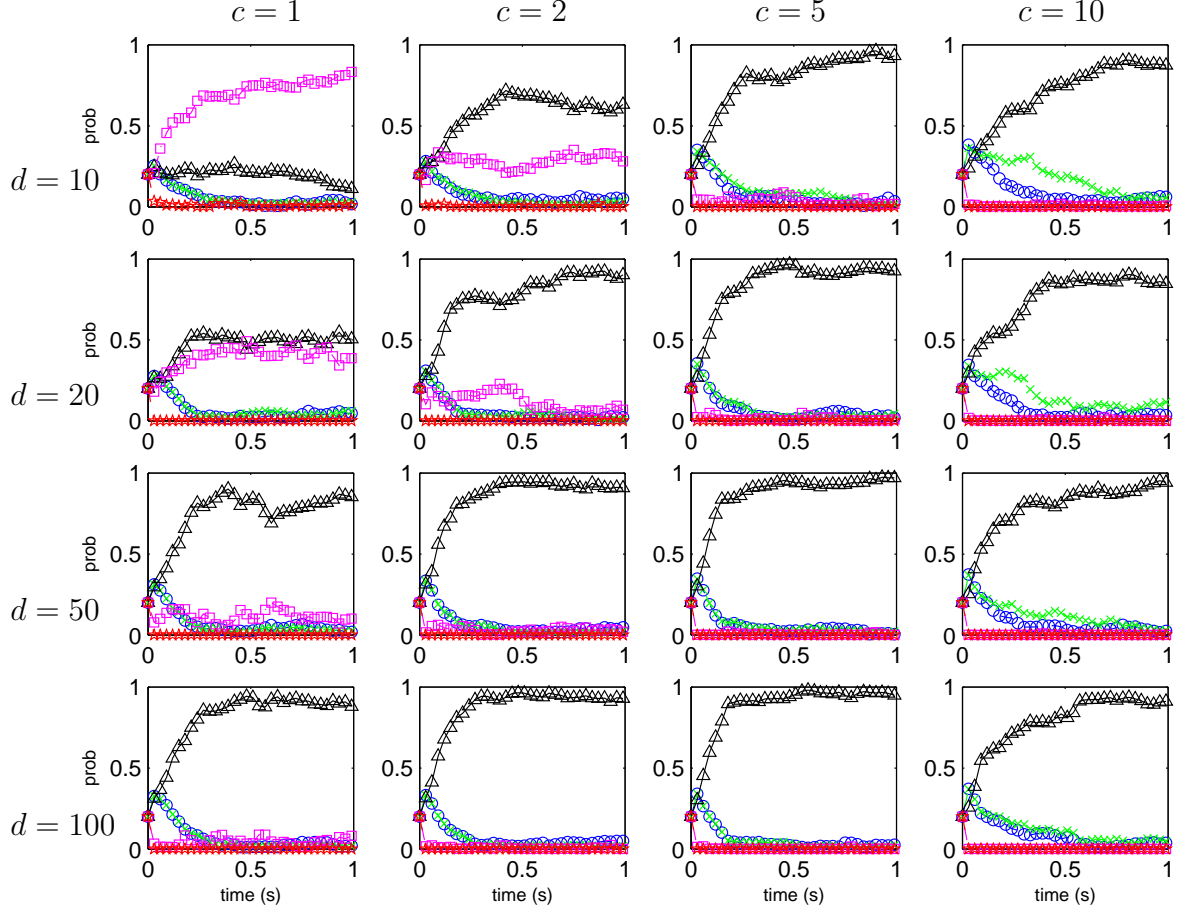


Figure 5.4 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 30$, $Q_{\text{true}} = 20$, $R_{\text{true}} = 1$. Filter 1 (\circ): $Q_1 = 20c/d^2$, Filter 2 (\times): $Q_2 = 20c/d$, Filter 3 (\triangle): $Q_3 = 20c$, Filter 4 (\square): $Q_4 = 20cd$, Filter 5 (\star): $Q_5 = 20cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

shown in Figures 5.1 through 5.9, we have taken for granted that we have accurately accounted for all of the non- Q model parameter values, while in Figures 5.10 and 5.11, we have intentionally based the bank of filters on an incorrect value for the true R . Finally, by choosing a higher-order model, we can regain some of the fidelity that we traded off earlier.

In general, as the discretization of the parameter space becomes coarser (i.e., as d increases), the elemental filter based upon the most correct model receives a larger

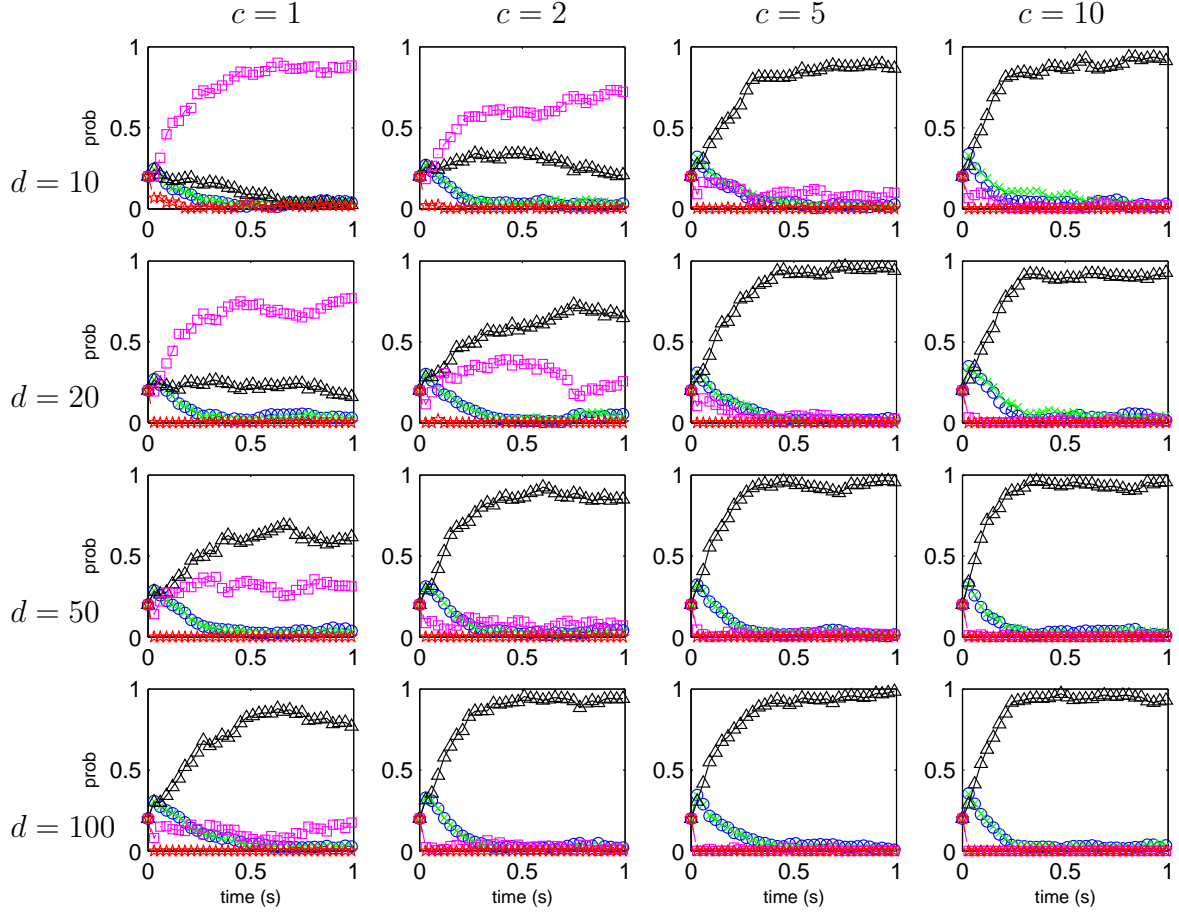


Figure 5.5 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 30$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 1$. Filter 1 (\circ): $Q_1 = 10c/d^2$, Filter 2 (\times): $Q_2 = 10c/d$, Filter 3 (\triangle): $Q_3 = 10c$, Filter 4 (\square): $Q_4 = 10cd$, Filter 5 (\star): $Q_5 = 10cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

share of the probability, this trend can be seen repeatedly in nearly all of the plots in Figures 5.1 through 5.11. A particularly good example of this trend can be seen in the far left (or first) column of plots in Figure 5.4 on page 5-18; the third filter (represented by the triangles) is based on the true values of the parameters and, as we can see, it gets a larger probability as we increase the coarseness of the discretization down the column from $d = 10$ to 20 to 50, and finally $d = 100$. In the $(c, d) = (1, 20)$ plot, the third and fourth (represented by the squares) elemental filters are virtually

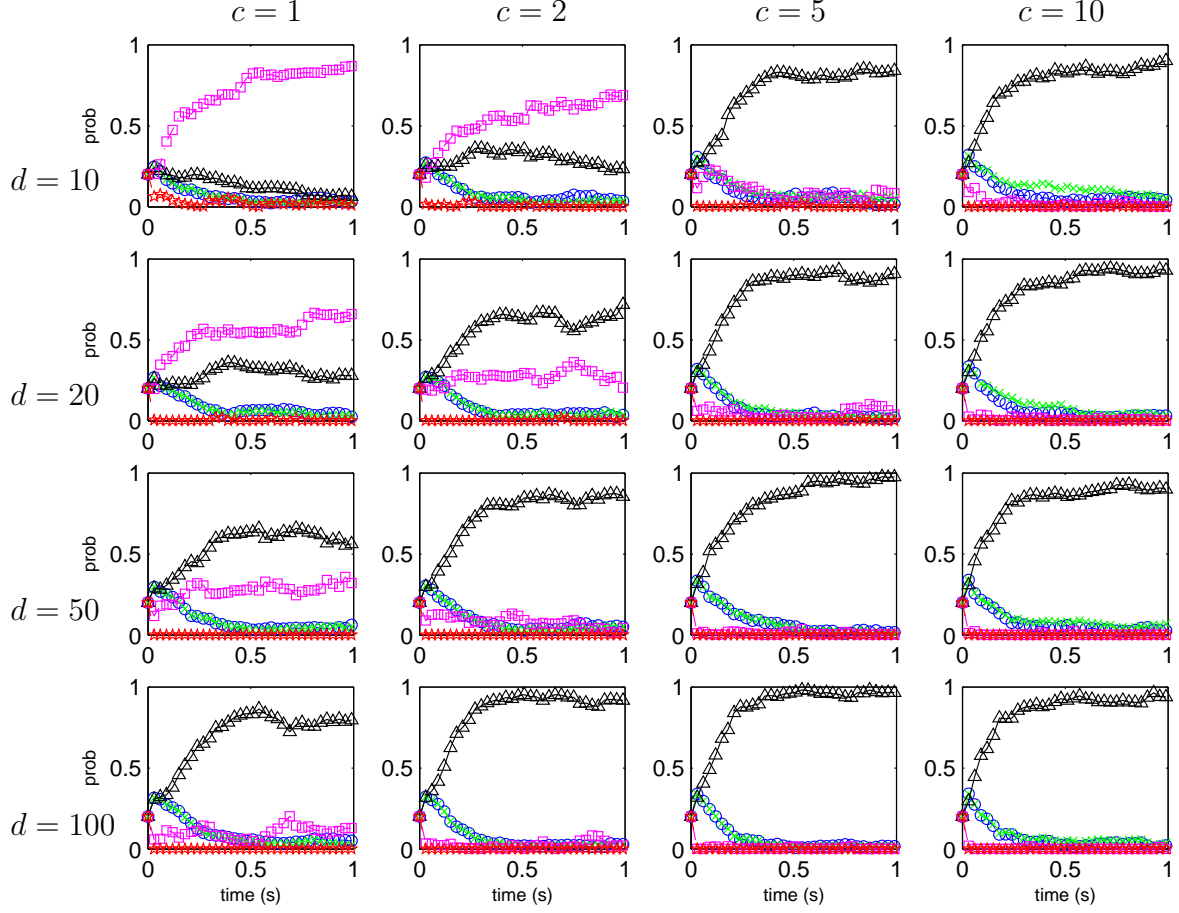


Figure 5.6 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 40$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 1$. Filter 1 (\circ): $Q_1 = 10c/d^2$, Filter 2 (\times): $Q_2 = 10c/d$, Filter 3 (Δ): $Q_3 = 10c$, Filter 4 (\square): $Q_4 = 10cd$, Filter 5 (\star): $Q_5 = 10cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

indistinguishable. Increasing the discretization to the $d = 50$ level, the third filter gains the clear majority of the probability. So, while the distinguishability of the elemental filters appears to degrade before it eventually improves, the probability flow to the elemental filter based on the most correct model clearly increases with an increasingly coarse level of discretization. Hence, as the discretization of the parameter space becomes coarser, the elemental filters become more properly distinguishable as the most appropriately modeled filter gains a larger share of the probability.

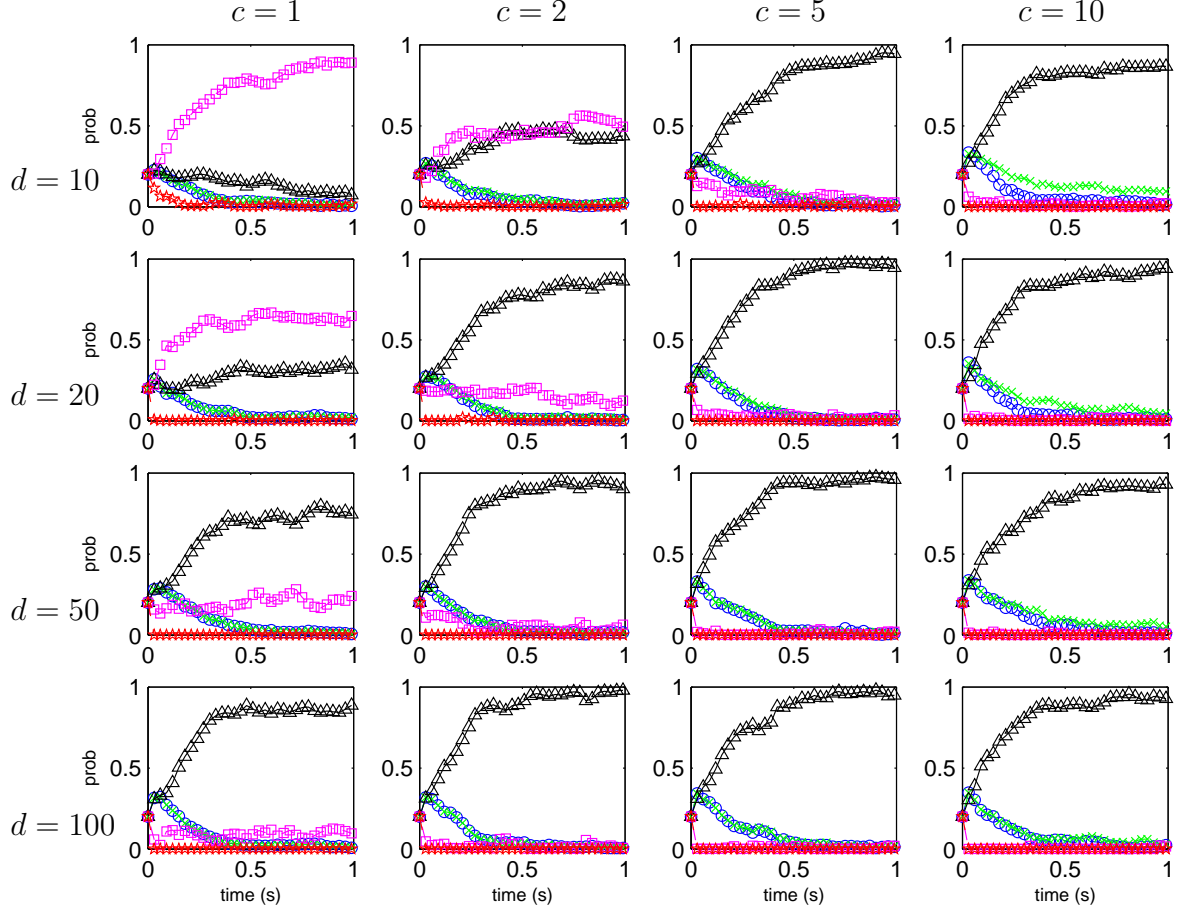


Figure 5.7 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 50$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 1$. Filter 1 (\circ): $Q_1 = 10c/d^2$, Filter 2 (\times): $Q_2 = 10c/d$, Filter 3 (\triangle): $Q_3 = 10c$, Filter 4 (\square): $Q_4 = 10cd$, Filter 5 (\star): $Q_5 = 10cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

The relative ratio of Q/R can be seen to influence the behavior of the bank of filters directly, as seen in these probability plots. When the ratio Q/R is very high, see Figures 5.1 and 5.2, the elemental filter based on the most correct model routinely gets the bulk of the probability. As we can see, the arrays of plots in these two figures are nearly identical; this is because they have the same Q/R ratio. On the other hand, halving the ratio Q/R from 100 to 50 for the cases shown in Figures 5.2 and 5.3 shows quite a different result for the “finely” discretized $d = 10$

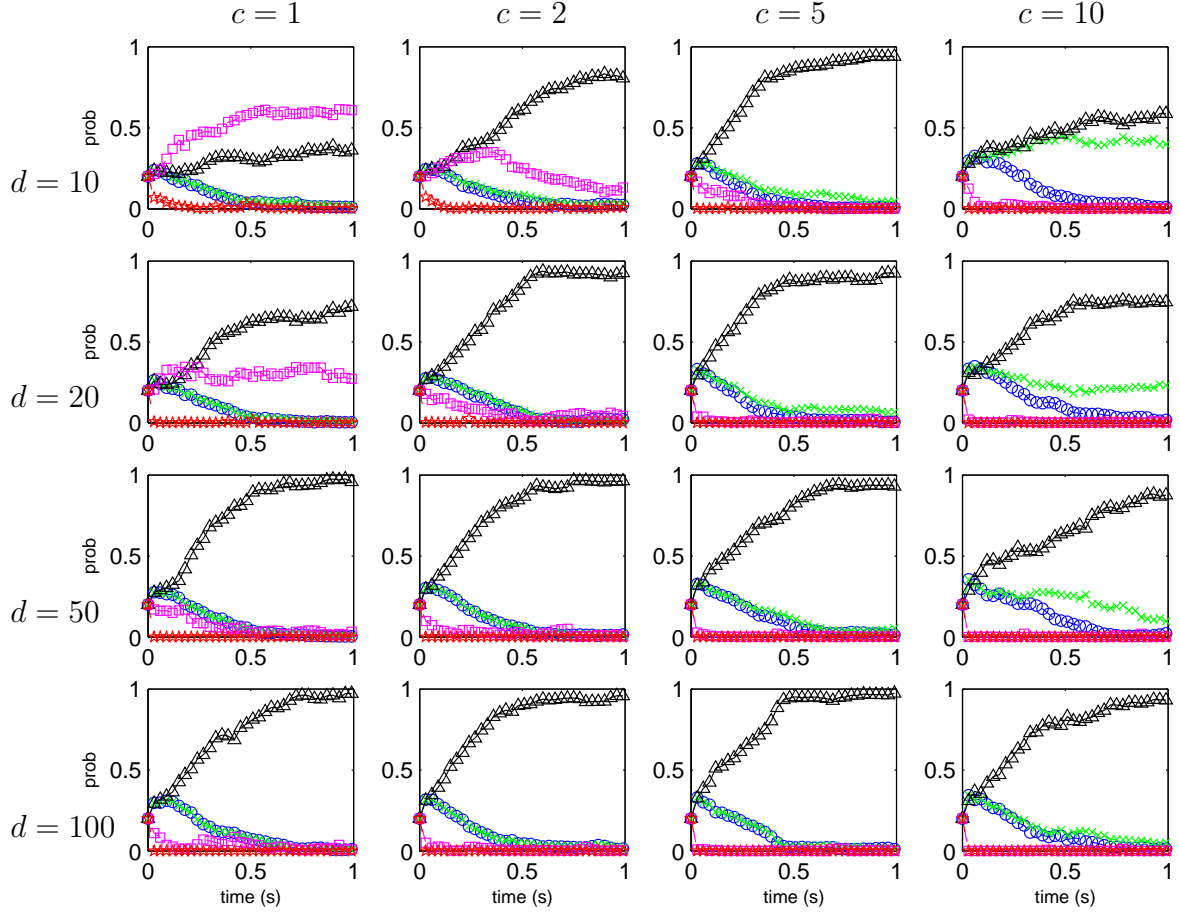


Figure 5.8 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 60$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 1$. Filter 1 (\circ): $Q_1 = 10c/d^2$, Filter 2 (\times): $Q_2 = 10c/d$, Filter 3 (\triangle): $Q_3 = 10c$, Filter 4 (\square): $Q_4 = 10cd$, Filter 5 (\star): $Q_5 = 10cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

top row of plots. Specifically, the third (second) and fourth (third) elemental filters are approaching indistinguishability in the centered $c = 1$ ($c = 10$ far right column) case. Thus, in terms of probability flow, the relative ratio of Q/R matters more than the particular values for Q and R . The ratio is important since Q/R dictates the steady state Kalman gain for each elemental filter — the steady state Kalman gain is directly proportional to the noise strength Q and inversely related to the noise covariance R .

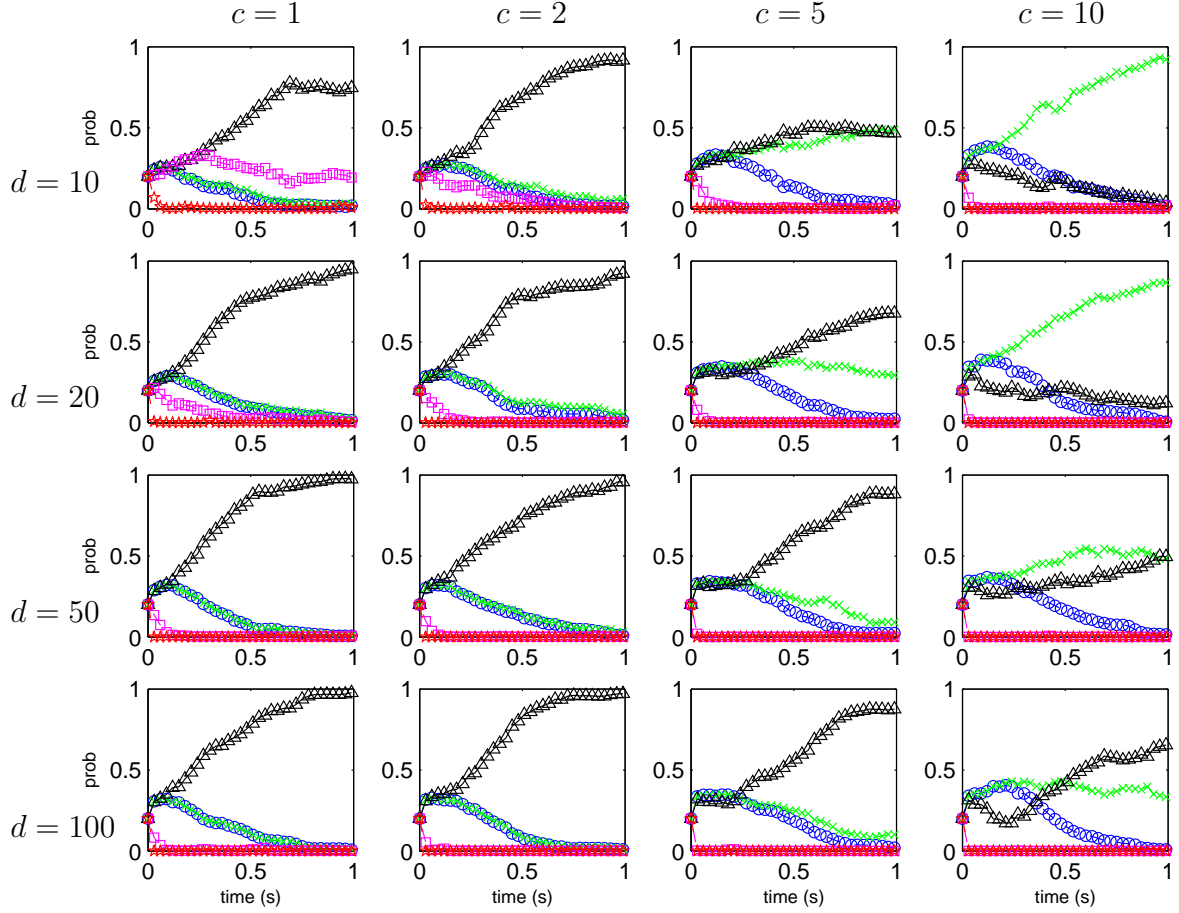


Figure 5.9 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 70$, $Q_{\text{true}} = 10$, $R_{\text{true}} = 1$. Filter 1 (\circ): $Q_1 = 10c/d^2$, Filter 2 (\times): $Q_2 = 10c/d$, Filter 3 (\triangle): $Q_3 = 10c$, Filter 4 (\square): $Q_4 = 10cd$, Filter 5 (\star): $Q_5 = 10cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

Assuming we have chosen R wisely, then it follows that when Q/R is low, we have chosen a value too low for our assumed Q . Hence, “overestimating Q ” at this point may well lead us to a value for Q that is about right and thus this elemental filter receives the majority of the probability for a filter bank constructed with a relatively fine discretization level; see for example, the fourth elemental filter (represented by the squares) in the top left plot in Figure 5.4. Scrolling down the first column of plots shows that, as the discretization becomes coarser, the “correct” filter

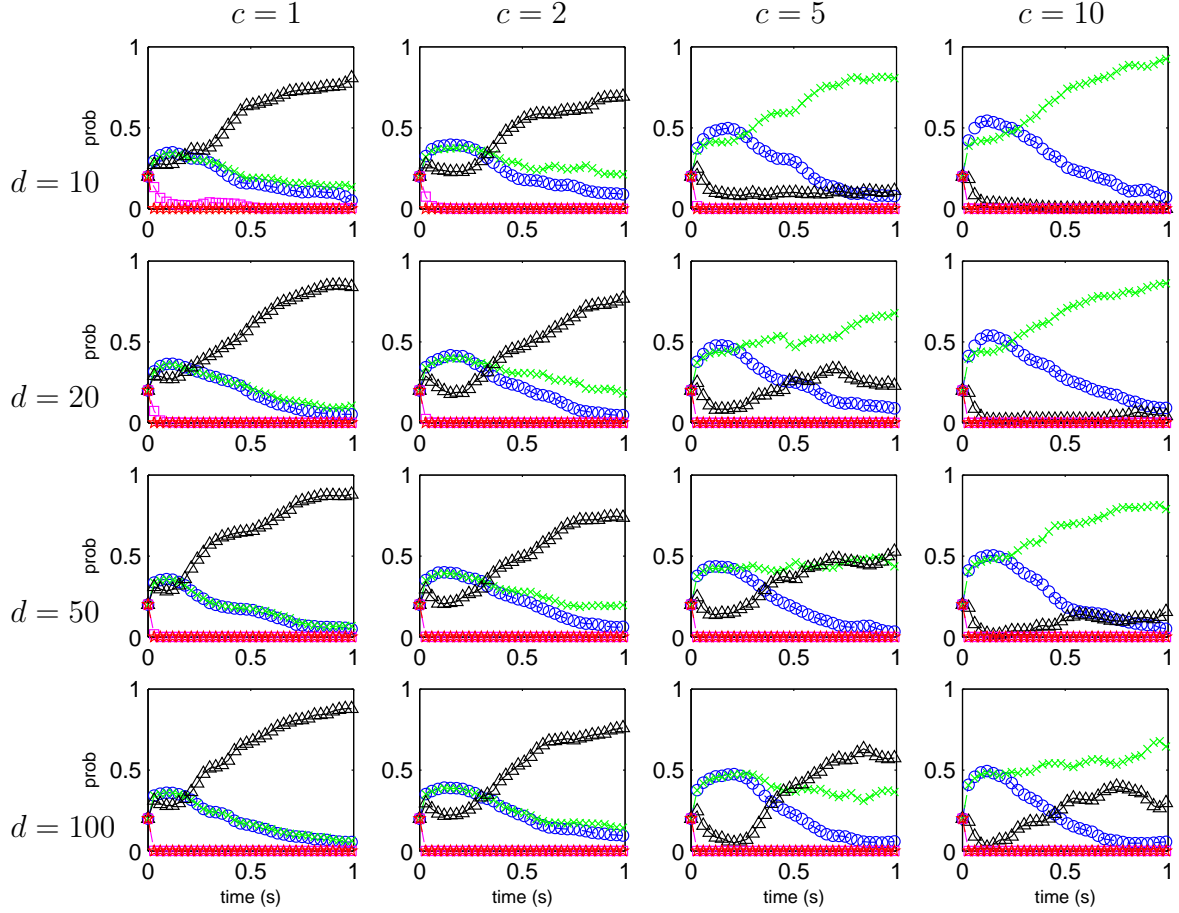


Figure 5.10 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 30$, $Q_{\text{true}} = 100$, $R_{\text{true}} = 1$, $R_{\text{filter}} = 10$. Filter 1 (\circ): $Q_1 = 100c/d^2$, Filter 2 (\times): $Q_2 = 100c/d$, Filter 3 (\triangle): $Q_3 = 100c$, Filter 4 (\square): $Q_4 = 100cd$, Filter 5 (\star): $Q_5 = 100cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

(the elemental filter based on the most correct hypothesis which is here indicated by the triangles since it is the third filter in the filter bank) receives an increasingly larger share of the probability, while the filter featuring an overestimate for Q (squares) receives a correspondingly smaller share of the probability. Looking at the second plot $(c, d) = (1, 20)$, we see that the third and fourth elemental filters are virtually indistinguishable as has been previously noted. As the discretization coarsens further to $d = 50$ in the third plot of the first column, the filters are once

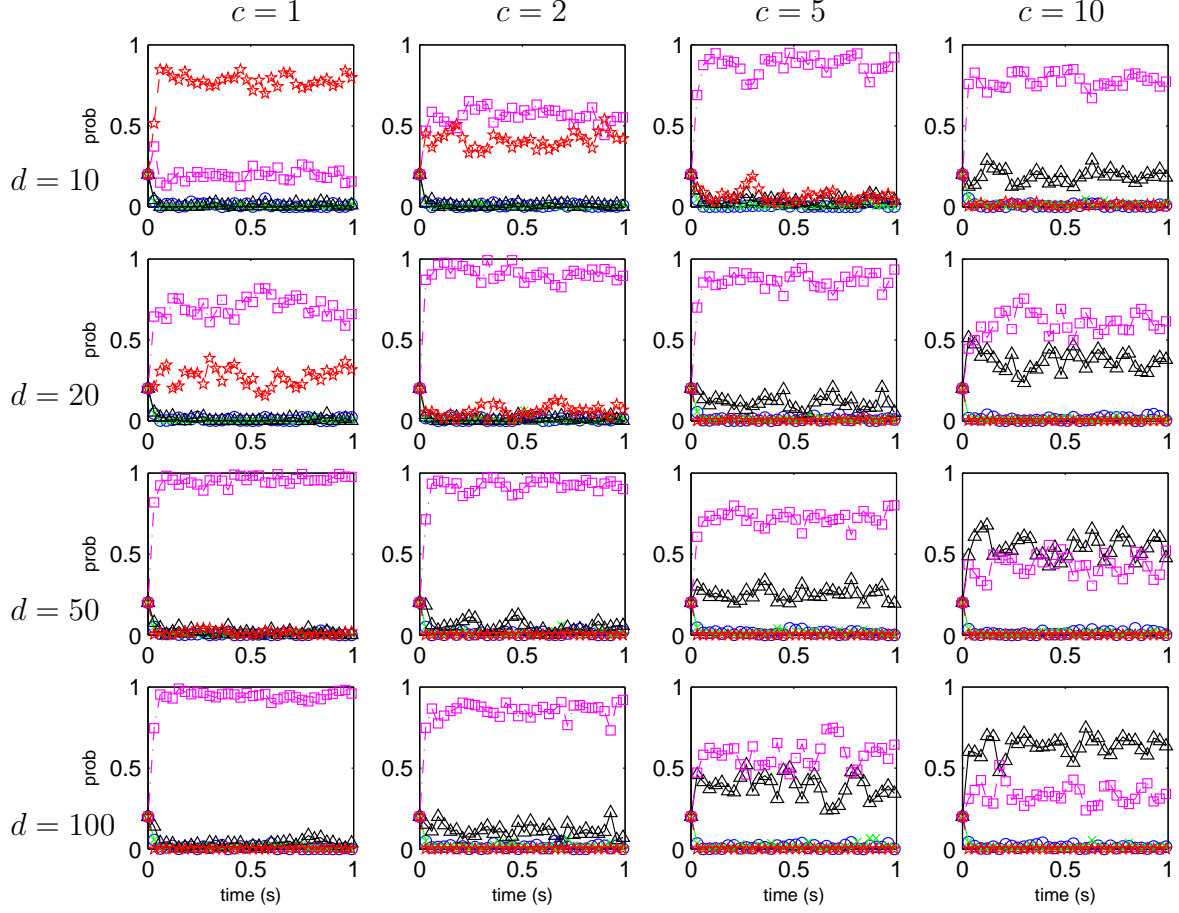


Figure 5.11 Simulation 1: Filter bank composition experiment. Hypothesis conditional probability histories for 16 cases of interest for $N = 30$, $Q_{\text{true}} = 100$, $R_{\text{true}} = 10$, $R_{\text{filter}} = 1$. Filter 1 (\circ): $Q_1 = 100c/d^2$, Filter 2 (\times): $Q_2 = 100c/d$, Filter 3 (\triangle): $Q_3 = 100c$, Filter 4 (\square): $Q_4 = 100cd$, Filter 5 (\star): $Q_5 = 100cd^2$. To make these small plots more legible, the mean hypothesis conditional probabilities are presented for only times $\{t_0, t_3, t_6, \dots, t_{99}\}$ versus all of $\mathbb{T} = \{t_0, t_1, t_2, \dots, t_{100}\}$.

again distinguishable; the third filter, constructed using the true value for Q , gathers the majority of the probability. This sequence of events occurs several times, as seen in Figures 5.4 and 5.5.

In the second column of plots, the “best” elemental filter is still the third filter (triangles); it was designed using a dynamics noise strength of twice Q_{true} . For the cases displayed in Figures 5.1 to 5.3, the ratio Q/R is relatively large and the best filter is chosen every time. As Q/R is reduced further, as in Figures 5.4 and 5.5, we

see that the $d = 10$ discretization level is too fine and results in probability being shared with an elemental filter featuring an even larger assumed value for Q in both figures, while in Figure 5.5, the fourth filter appears to be the best match. For both cases, as the discretization is made coarser by increasing d , the hypothesis put forth by the third filter gathers the highest probability in the filter bank.

We have seen in Figures 5.1 to 5.5 how the Q/R ratio and the discretization of the dynamics noise strength, Q , have generally affected the probability flow to the elemental filters. Now, let's see how the distinguishability and probability flow is affected by the order of the model, N . In Figures 5.5 to 5.9, we have gradually increased the order of the model to the point where the third (most-correctly modeled) filter is once again clearly dominant — nearly as good as it was for the case when $Q/R = 100$ as seen in Figures 5.1 and 5.2. If you recall our discussion early in this chapter on the state transition matrix, it is quite remarkable to see the accumulation of effects contributed by the 31st to 70th “states”. Recall that $\left[\tilde{\Phi}\right]_{30}(t_{i+1} - t_i) = 2.6 \times 10^{-39}$, while another computation gives $\left[\tilde{\Phi}\right]_{70}(t_{i+1} - t_i) = 9.3 \times 10^{-211}$. Thus, the states corresponding to large N values, i.e., $N > 30$, which are nearly driven to zero during each propagation cycle, are still quite important!

In the previous comments, we have implicitly assumed that our models were based on the true measurement noise covariance. While overestimation of the true measurement noise covariance, as seen in Figure 5.10, slows the probability flow to the elemental filter based on the correct model, underestimating the covariance, as seen in Figure 5.11 “causes” the filters to attribute the increased (and unexpected) errors to the dynamics noise strength. The real problem is that the state estimation suffers in both cases, devastatingly for the underestimation case.

Now that we have graphically seen how the discretization level, d , the relative quality of the dynamics and measurement models, Q/R ratio, and the order of the model, N , affect the distinguishability of the elemental filters, we shall investigate

the effects of filter initialization of the state estimate, \hat{x} , and the state covariance, P , in the next simulation.

5.3 *Simulation 2*

The first five simulations feature adaptation to an uncertain noise environment and have an initial state estimate \hat{x}_0 of 25 °C. However, Table 5.1 on page 5-3 gives the true value as 20 °C; hence there is a 5 °C bias in our initial state estimate. In a real-world scenario, we might not know that our initial state estimate was off by 5 °C; thus, it is entirely possible that we would set the initial state covariance, P_0 , too low. In Table 5.4, we have set the initial state covariance estimate, P_0 , equal to 25 (°C)² to account for the fact that we know that there is a bias on the order of 5 °C in our initial state estimate.

Our goal in this simulation is to improve state estimation performance by adapting to an unknown system dynamics noise strength; we assume (however imperfectly) that the other system parameters and noise statistics are completely known, with the exception that we consider what happens when we fail to set the initial state covariance, P_0 , properly. If the filter-assumed initial state covariance is too small for the assumed \hat{x}_0 in the true scenario versus the filter, then the filter will not properly adjust the gain and will weight the initial measurements too lightly. On the other hand, when P_0 is too large, the filter becomes too responsive to initial measurements and disregards the system dynamics which allow us to propagate the state estimate since the last measurement update. So, when Q is accurately estimated by the MMAE (and we have properly set the initial state covariance), then we obtain state estimation performance approaching that of a single Kalman filter with artificial knowledge of the correct parameter values.

The truth model parameters are given in Table 5.1. Some of the design parameters for the three elemental filters of the MMAE are included in Table 5.6. Note that Q_{true} is the true value of the dynamics noise strength as listed in Table 5.1.

Case	Filter	Q	R	\hat{x}_0	P_0
1	1	$\frac{1}{100}Q_{\text{true}} = \frac{1}{20}$	R_{true}	25	$P_0^{\text{poor}} = 1$
	2	$Q_{\text{true}} = 5$			
	3	$100Q_{\text{true}} = 500$			
2	1	$\frac{1}{100}Q_{\text{true}} = \frac{1}{20}$	R_{true}	25	$P_0^{\text{good}} = 25$
	2	$Q_{\text{true}} = 5$			
	3	$100Q_{\text{true}} = 500$			
3	1	$\frac{1}{10}Q_{\text{true}} = \frac{1}{2}$	R_{true}	25	$P_0^{\text{poor}} = 1$
	2	$10Q_{\text{true}} = 50$			
	3	$1000Q_{\text{true}} = 5000$			
4	1	$\frac{1}{10}Q_{\text{true}} = \frac{1}{2}$	R_{true}	25	$P_0^{\text{good}} = 25$
	2	$10Q_{\text{true}} = 50$			
	3	$1000Q_{\text{true}} = 5000$			

Table 5.6 Simulation 2: Elemental filter parameters for the initial state error covariance experiment. Cases 1 and 2 feature a filter bank centered on the true value of Q , while Cases 3 and 4 are for an arrangement that overestimates the true value of Q (centered on $10Q_{\text{true}}$ rather than Q_{true} itself). Note that the assumed \hat{x}_0 value is 5 °C greater than the true x_0 of 20 °C.

Since the true value for \hat{x}_0 is 20 °C versus the 25 °C assumed for the filters in this simulation, a poor choice for the initial state covariance P_0 would be $P_0^{\text{poor}} = 1$, while a good choice would be $P_0^{\text{good}} = 25$. The poor choice, $P_0^{\text{poor}} = 1$, reflects a *poor* assessment of the initial state estimate bias: the actual value of x_0 is at a 5σ point according to the filter-assumed P_0 . An undersized initial state covariance degrades the filter’s ability to converge on a good estimate because it essentially directs the filter to underweight the measurement updates by keeping the gain low. Thus, the responsiveness of the elemental filter is inhibited by a poor choice for the initial state covariance. The filter can generally recover from this error, however, it takes several propagate/update cycles for this state error to settle out. A good choice for the initial state covariance ($P_0^{\text{good}} = 25$) accurately portrays the size of the initial state estimate bias.

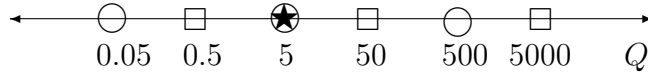


Figure 5.12 Simulation 2: System dynamics noise strength. Legend: \bigcirc elemental filters for Cases 1 and 2; \square elemental filters for Cases 3 and 4; \star true parameter within the filter bank. (The filter spacing is nonlinear for illustration purposes.)

The relative spacing (on a log scale) for the assumed dynamics noise strength for the two sets (four cases) of three elemental filters is displayed in Figure 5.12. Additionally, successive members of the list of three filters are separated by two orders of magnitude in Q , have the same R , and the same initial temperature bias and state covariance for each case. The large separation is due to the indistinguishability of filters for closely spaced values of Q as shown in the previous experiment. Since the effects of the dynamics noise strength are ascertained using measurements, level of the measurement-corruption noise covariance has an impact on how well we can estimate Q . Consequently, for better measurements (“smaller” R), we can get a better fix on Q .

As expected, the elemental filter designed for a model that “slightly” overestimates the true noise strength, Q_{true} , matches the real world the best, as indicated by its high mean hypothesis conditional probability, p_2 , relative to the other elemental filters in the filter bank (p_1 and p_3); compare Figure 5.13(a) to (c) for the poor P_0 models and Figure 5.13(b) to (d) for the good P_0 models when the Q/R ratio is near unity as seen previously in simulation one. Recall that the hypothesis conditional probability, p_k , increases as the sequence of residuals have a filter-computed covariance that is most in consonance with the actual covariance of the measurement residuals⁷. In this example, elemental filter 2 is the most properly modeled filter and its probability tends towards one the strongest in Case 3, as seen in Figure

⁷It has been shown [94, 129] that the sequence of residuals $\{\mathbf{r}_k(t_i)\}$ resulting from linear filtering in additive noise forms a zero-mean white Gaussian sequence with known residual covariance $\mathbf{A}_k(t_i)$. Thus, if a filter model matches the “true” system, then the residual $\mathbf{r}_k(t_i)$ will be a zero-mean white Gaussian process with known residual covariance $\mathbf{A}_k(t_i)$.

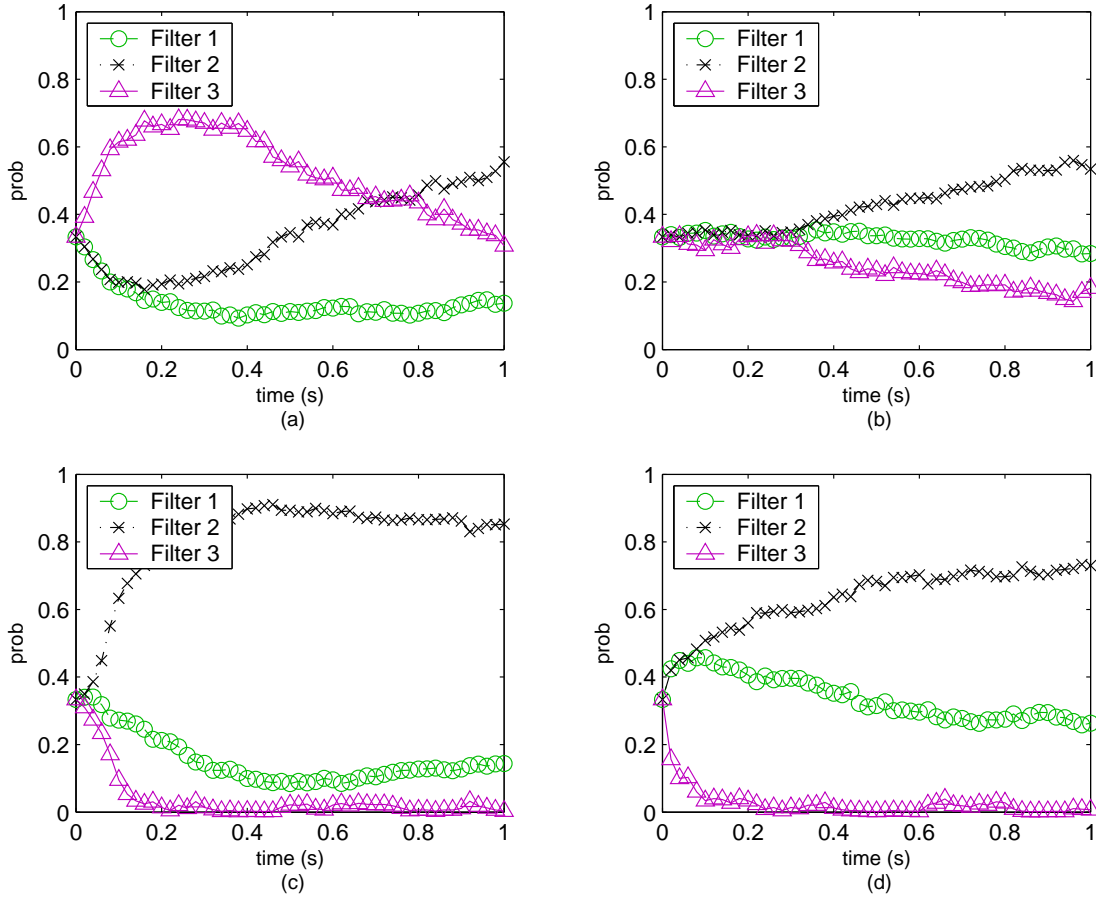


Figure 5.13 Simulation 2: Initial state covariance experiment — hypothesis conditional probability flow. (a) Case 1: P_0^{poor} , $\{\frac{1}{100}Q_{\text{true}}, Q_{\text{true}}, 100Q_{\text{true}}\}$. (b) Case 2: P_0^{good} , $\{\frac{1}{100}Q_{\text{true}}, Q_{\text{true}}, 100Q_{\text{true}}\}$. (c) Case 3: P_0^{poor} , $\{\frac{1}{10}Q_{\text{true}}, 10Q_{\text{true}}, 1000Q_{\text{true}}\}$. (d) Case 4: P_0^{good} , $\{\frac{1}{10}Q_{\text{true}}, 10Q_{\text{true}}, 1000Q_{\text{true}}\}$. To make these plots clearer, only the mean hypothesis conditional probabilities for times $\{t_0, t_2, \dots, t_{100}\}$ are displayed.

5.13(c), while, at the same time, the other elemental filter hypothesis conditional probabilities tend toward zero. The good probability flow for Case 3 is due to a combination of circumstances. An elemental filter that has “slightly” overestimated the dynamics noise strength can oftentimes compensate for an inadequate initialization of the state covariance. On the other hand, an elemental filter with too low of an assumed Q is quickly recognized by the MMAE as a mismodeled filter since its rate of convergence is not increased by a large assumption for Q . While a poor setting for P_0 degrades state estimation performance, this may have a beneficial impact on

parameter estimation when the best elemental filter design is one that assumes a slightly too large value for the parameter of interest, i.e., for the case when we do not have an elemental filter based on the true parameter value — this research did not address this possible “enhancement” any further.

The relatively poor distinguishability of the trio of filters in plots (a) and (b) of Figure 5.13 is due, in part, to the low Q/R ratio — this effect was seen previously in the far left column of plots in Figure 5.5. The probability flow in Case 2 is better than for Case 1 because of a better assumption for P_0 . The poor P_0 in Case 1 serves to obscure the best elemental filter initially because the too large assumption for Q combined with the too small P_0 for elemental filter 3 yields what appears to the MMAE as a good assessment of the error covariance; however, elemental filter 2 eventually absorbs the probability initially given to elemental filter 3 as it “flushes” out the poor initialization of the error covariance given by P_0 .

On the other hand, the elemental filters shown in plots (c) and (d) of Figure 5.13, are more distinguishable because the best match is a filter based on an overestimate for the dynamics noise strength; the far right column of plots in Figure 5.5 is a good example of how the overestimate of the noise strength can lead to increased distinguishability between filters.

Now let’s look a little closer at general trends evident for elemental filter 1. When the (simulated) real-world noise strength exceeds the hypothesized noise strength, the filter’s residuals look bad and consequently, the likelihood quotient $\mathbf{r}_1^T(t_i) \mathbf{A}_1^{-1}(t_i) \mathbf{r}_1(t_i)$ grows larger; Equation (5.6) informs us that we should expect the likelihood quotient, $E\{\mathbf{L}_1(t_i)\} = \text{tr}\{\mathbf{A}_1^{-1}(t_i) \mathbf{A}_{\text{true}}(t_i)\}$, to grow larger as the true residual covariance “increases” relative to the assumed filter-computed residual covariance. Therefore, the probability that the elemental filter is based on a good model which accurately reflects the real-world, decreases. In all four plots of Figure 5.13, elemental filter 1 is clearly not based on the best model; this is most evident

in plots (a) and (c), while in plots (b) and (d), the higher quality, “good”, estimate for P_0 somewhat masks the much too low value for Q .

When the filter bank is centered on the true dynamics noise strength, see plots (a) and (b) in Figure 5.13, the low Q/R ratio generally produces marginally distinguishable filters. Additionally, the quality of the initial state covariance estimate, P_0 , strongly influences the initial performance of the filters. The quality of the initial estimate determines how long it takes the filter to recover from a poor estimate. As the state covariance converges to its true value, the second elemental filter absorbs the majority of the probability. At the 1-second mark, elemental filter 2 for both cases, shown in plots (a) and (b), gain roughly three fifths of the probability. In plot (a), the third elemental filter initially dominates because the poor (underestimate of) P_0 nicely balances the overestimate of the dynamics noise strength. This unfortunate effect can be seen clearly by rewriting the filter-computed residual covariance: $\mathbf{A}_k(t_i) = \mathbf{H}_k(t_i) \mathbf{P}_k(t_i^-) \mathbf{H}_k^T(t_i) + \mathbf{R}_k(t_i)$ using the expression for the propagated state covariance to give (without the k subscripts)

$$\mathbf{A}(t_i) = \mathbf{H}(t_i) [\Phi(t_i, t_{i-1}) \mathbf{P}(t_{i-1}^+) \Phi^T(t_i, t_{i-1}) + \mathbf{Q}_d(t_{i-1})] \mathbf{H}^T(t_i) + \mathbf{R}(t_i) \quad (5.11)$$

In plot (b), the low Q/R ratio combined with a good P_0 estimate create marginally distinguishable filters, which eventually become more distinguishable and the properly modeled second elemental filter gains the majority of the probability.

In plots (c) and (d) of Figure 5.13, we show the results for the case when the filter bank does not contain an elemental filter that matches the true values assumed by the simulation. In this case, the elemental filter which looks the best, i.e., the elemental filter with filter-computed residual covariance that is most in consonance with the true residual covariance, is the one which slightly overestimates the dynamics noise strength. To illustrate, consider three elemental filters, numbered 1, 2, and 3, based on higher Q values as the index number increases. For elemental

filter 1 based on a too low Q value, the likelihood quotient $\mathbf{r}_1^T(t_i) \mathbf{A}_1^{-1}(t_i) \mathbf{r}_1(t_i)$ will grow much larger than m (because the measurement residuals, $\mathbf{r}_1(t_i)$, are much larger than anticipated). On the other hand, for elemental filter 3 based on too large a Q value, $\mathbf{r}_3^T(t_i) \mathbf{A}_3^{-1}(t_i) \mathbf{r}_3(t_i)$ won't be as large because the filter-computed covariance $\mathbf{A}_3(t_i)$ is so much larger. If the true value $\mathbf{A}_{\text{true}}(t_i)$ increases over time (in a ramp fashion, for example), then when $\mathbf{A}_{\text{true}}(t_i) = \mathbf{A}_2(t_i)$, elemental filter 2 should absorb most of the probability. However, when $\mathbf{A}_{\text{true}}(t_i)$ gets larger than $\mathbf{A}_2(t_i)$ (even by a small amount), $\mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)$ can become significantly larger than m , with the result that the probability flows to the elemental filter 3, even if $\mathbf{A}_3(t_i)$ is much larger than $\mathbf{A}_{\text{true}}(t_i)$.

If a chosen discretization is so coarse, that this phenomenon causes estimation problems, then use a finer discretization — for a finer discretization, the “next higher” elemental filter that absorbs the probability will not be based on a much-too-high value of Q or R . Furthermore, specifically consider a moving-bank MMAE (discussed in Section 2.6) to allow for a fine discretization without the burden of populating the filter bank with an excessively large number of elemental filters that results from the requirement to cover the entire range of possible Q or R values.

While elemental filter 3 is initially favored in Case 1, as shown in plot (a), it is rejected the quickest in Case 4, see plot (d), as compared to the other cases, because (1) the too large assumption of the dynamics noise strength is not obscured by a poor initial state covariance estimate as in Case 3 and, (2) compared to Case 2, elemental filter 3 assumes a much too large value for Q . The key to good performance is to have an elemental filter based on a model that only *slightly* overestimates the noise strength; when the assumed Q value is significantly too high, the filter becomes overly responsive to the measurements and then the subset of filters which feature overestimated noise strengths become less distinguishable.

When comparing the probability gathered by elemental filter 1 in plots (c) and (d), we see that elemental filter 1 is rejected more quickly in Case 3. The poor

initialization of the error covariance, P_0 , artificially helps us to see that the too low assumption for Q is completely inadequate to model the error covariance given the measurements taken. On the other hand, the more appropriate P_0 assigned in Case 4 does not give the too low Q assumption that extra boost to enhance the probability flow away from elemental filter 1.

Now that we have compared and contrasted many aspects of these four cases, we shall look at the individual cases more closely. The remainder of this section is comprised of alternating discussions and compilations of figures which apply to the specific cases.

Case 1: A poor initial state covariance and a set of parameter values for the bank of elemental filters that is centered on the true dynamics noise strength. A close inspection of Figure 5.14(a - f) for elemental filter 1 shows that the state estimate is slowly converging towards the true state — slowly because we have told the filter that the dynamics model is very good (much better than it truly is) and thus not to put too much trust in the measurements. Additionally, this slowness is exacerbated by the poor choice for the initial state covariance, as can be easily demonstrated by comparison to the Case 2 results, i.e., see Figure 5.18(a - f). In particular, look at time zero in Figure 5.14(d, f) to observe that the state estimate mean error (solid line) and the mean plus and minus one sigma (dash-dot line) values are outside the zero plus and minus one filter-computed sigma bounds (gray dashed line) initially created by the initial state covariance. For proper operation, the filter must operate inside of these bounds. To see how the estimate progresses along the entire rod, see Figure 5.14(i - p) on the continuation page of the figure; these plots also list the RMS error for each displayed time instant. By comparison, for elemental filter 2 in Figure 5.15(a - f), we see that the state estimate converges more quickly because we have a larger and more realistic assumed value of the noise strength. Continuing this theme for elemental filter 3, we see in Figure 5.16(a - f) that it converges very rapidly and then overshoots and continually overreacts to the noise

entering through the measurements. This can be quantified by examining the Q/R ratio, which determines the filter gain, K . As Q/R increases, so does the Kalman gain, which results in a wider filter bandwidth — faster responsiveness. On the other hand, lower Q/R produces a smaller Kalman gain and thus the system must wait for measurements to make substantial changes in its estimates. Additionally, elemental filter 3 absorbs the bulk of the probability early on because its higher assumed Q allows it to compensate for the poor choice for P_0 .

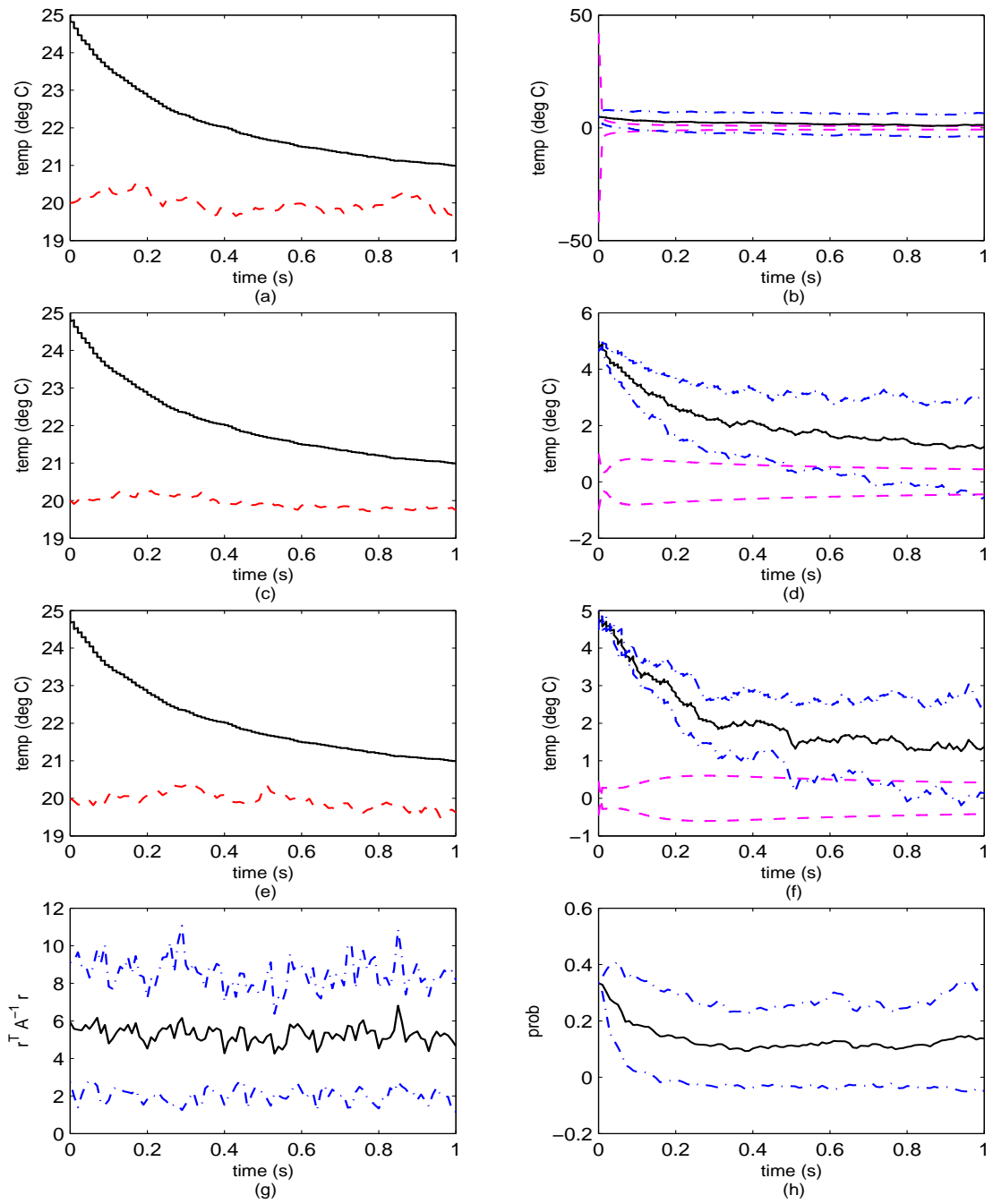


Figure 5.14 Simulation 2, Case 1 (P_0^{poor}): Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

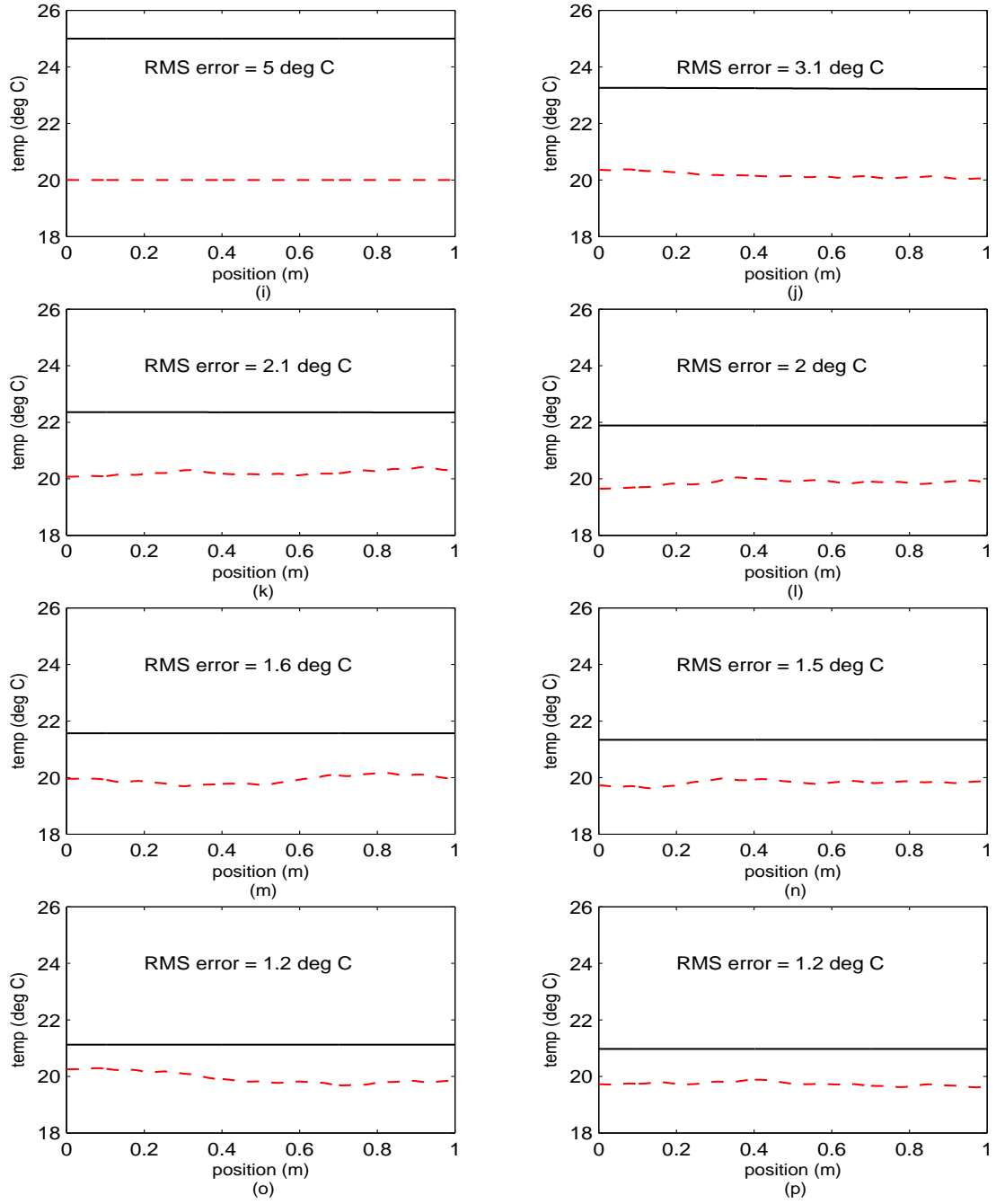


Figure 5.14 Simulation 2, Case 1 (P_0^{poor}): Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

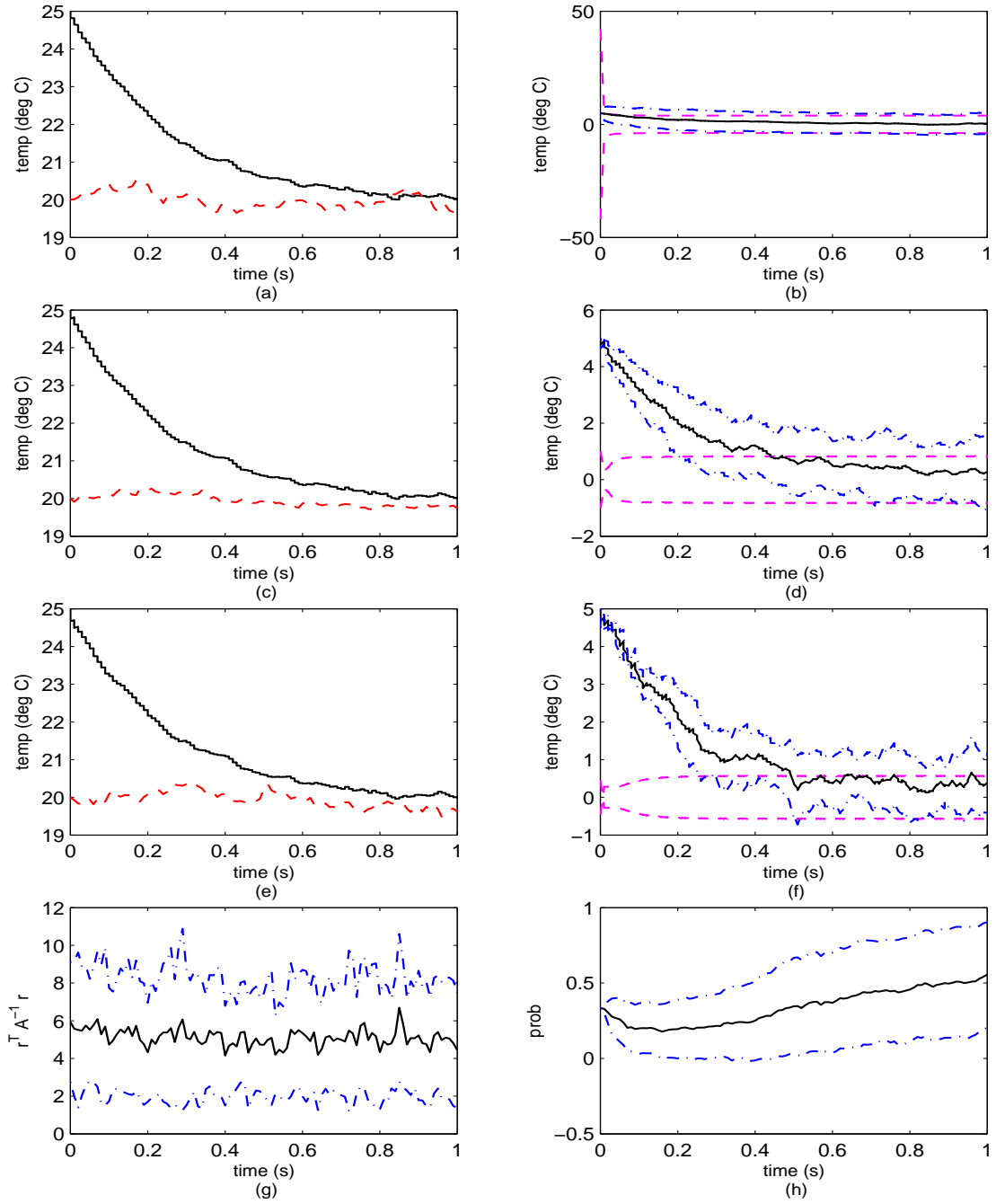


Figure 5.15 Simulation 2, Case 1 (P_0^{poor}): Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

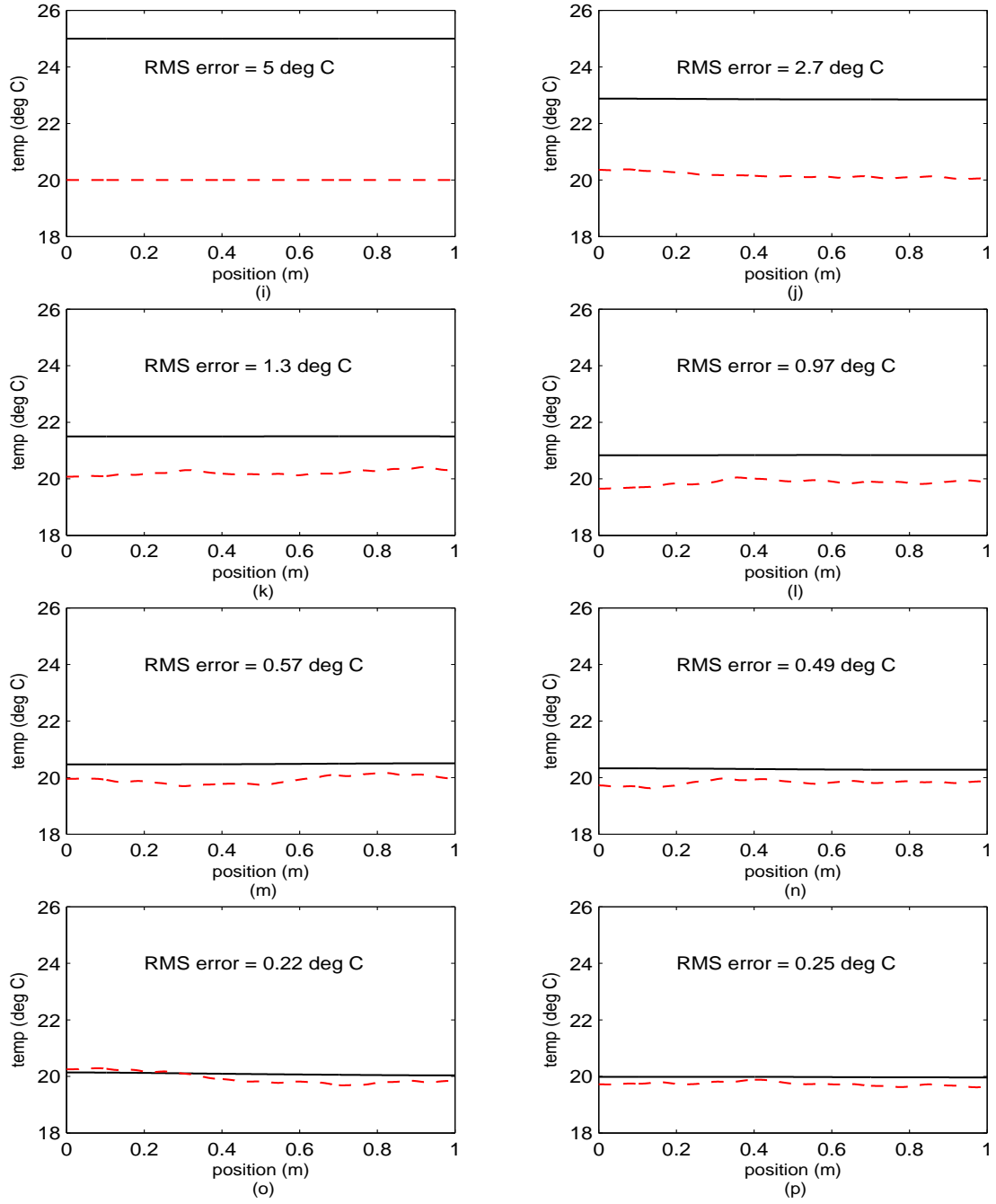


Figure 5.15 Simulation 2, Case 1 (P_0^{poor}): Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

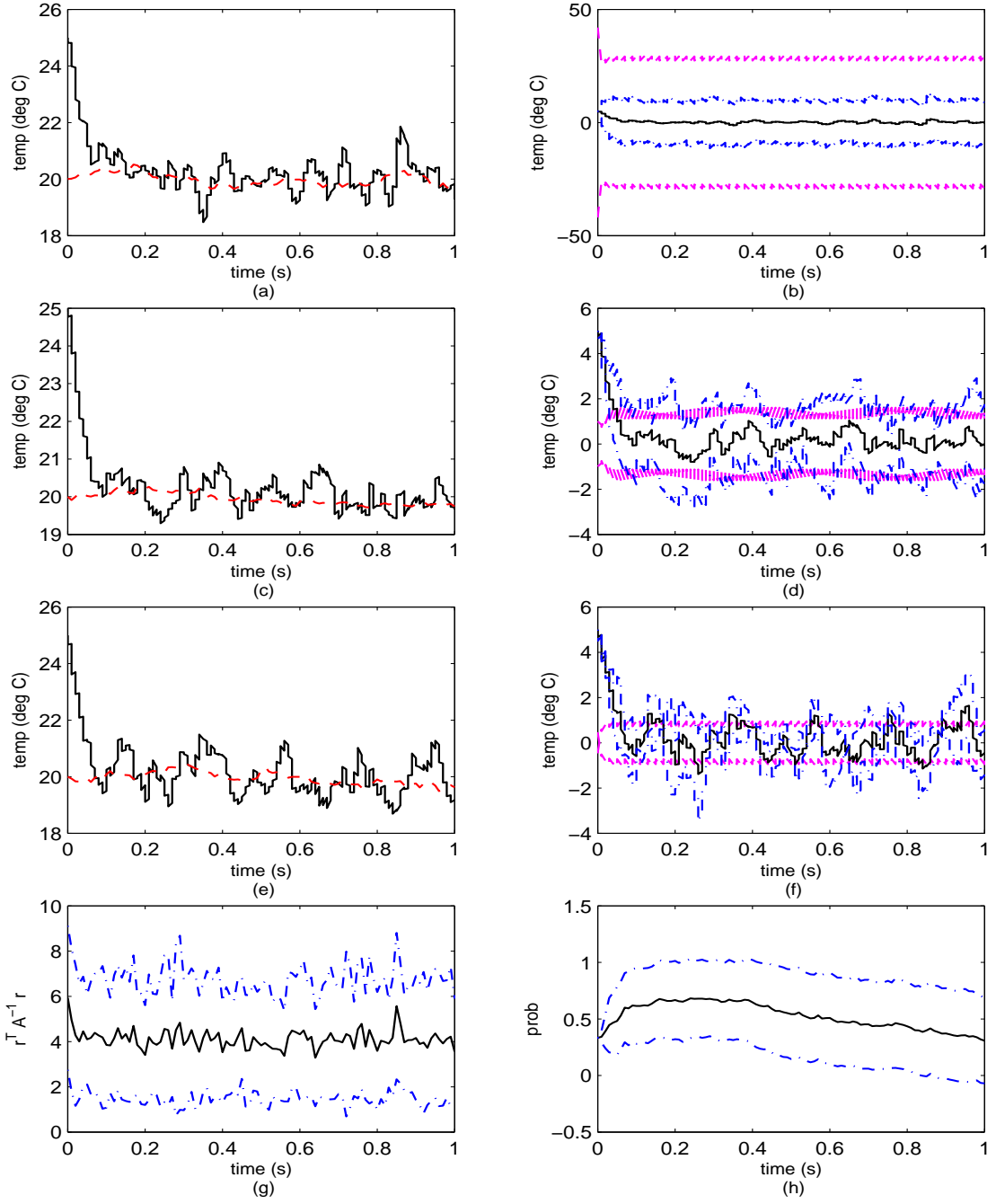


Figure 5.16 Simulation 2, Case 1 (P_0^{poor}): Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

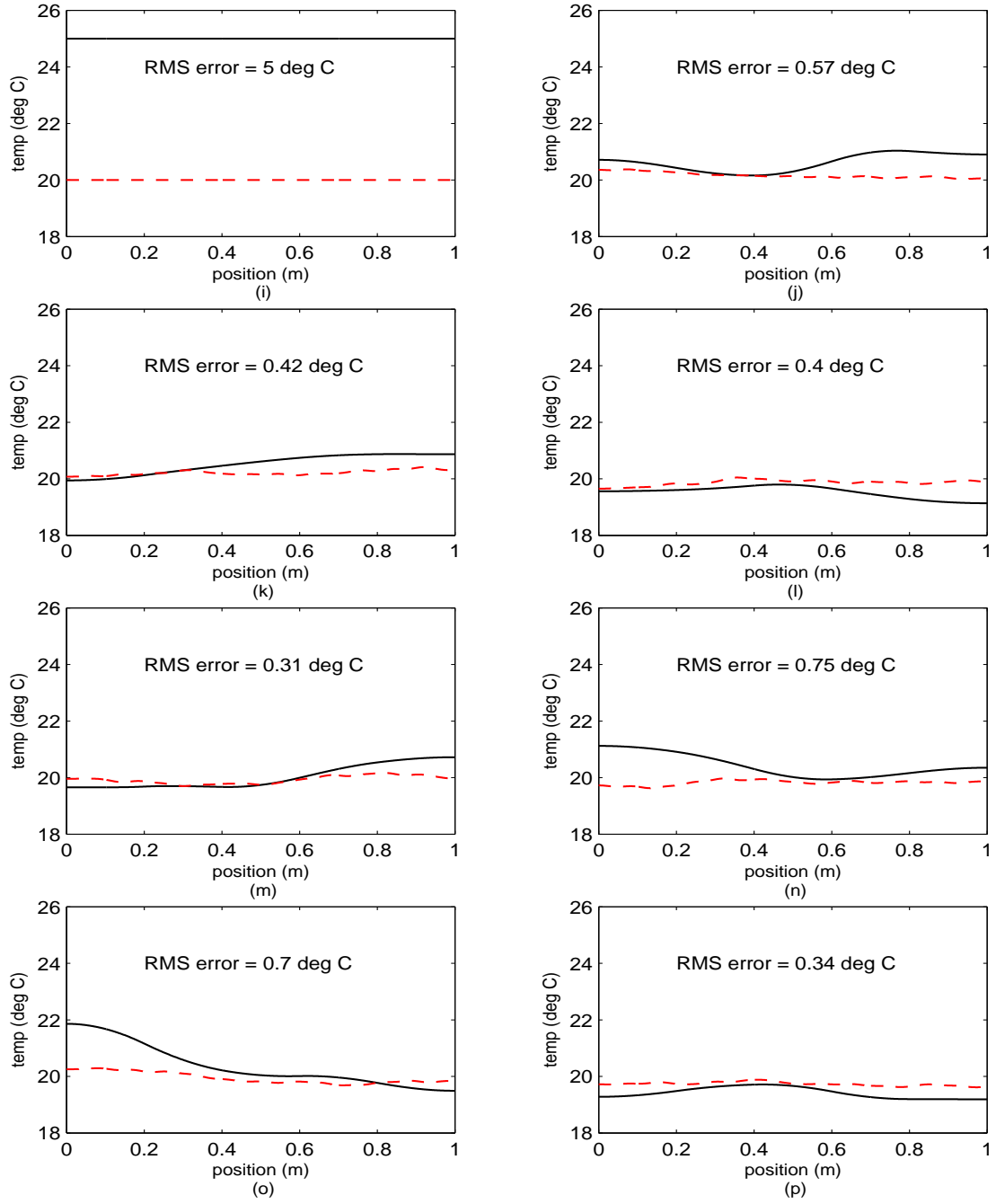


Figure 5.16 Simulation 2, Case 1 (P_0^{poor}): Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

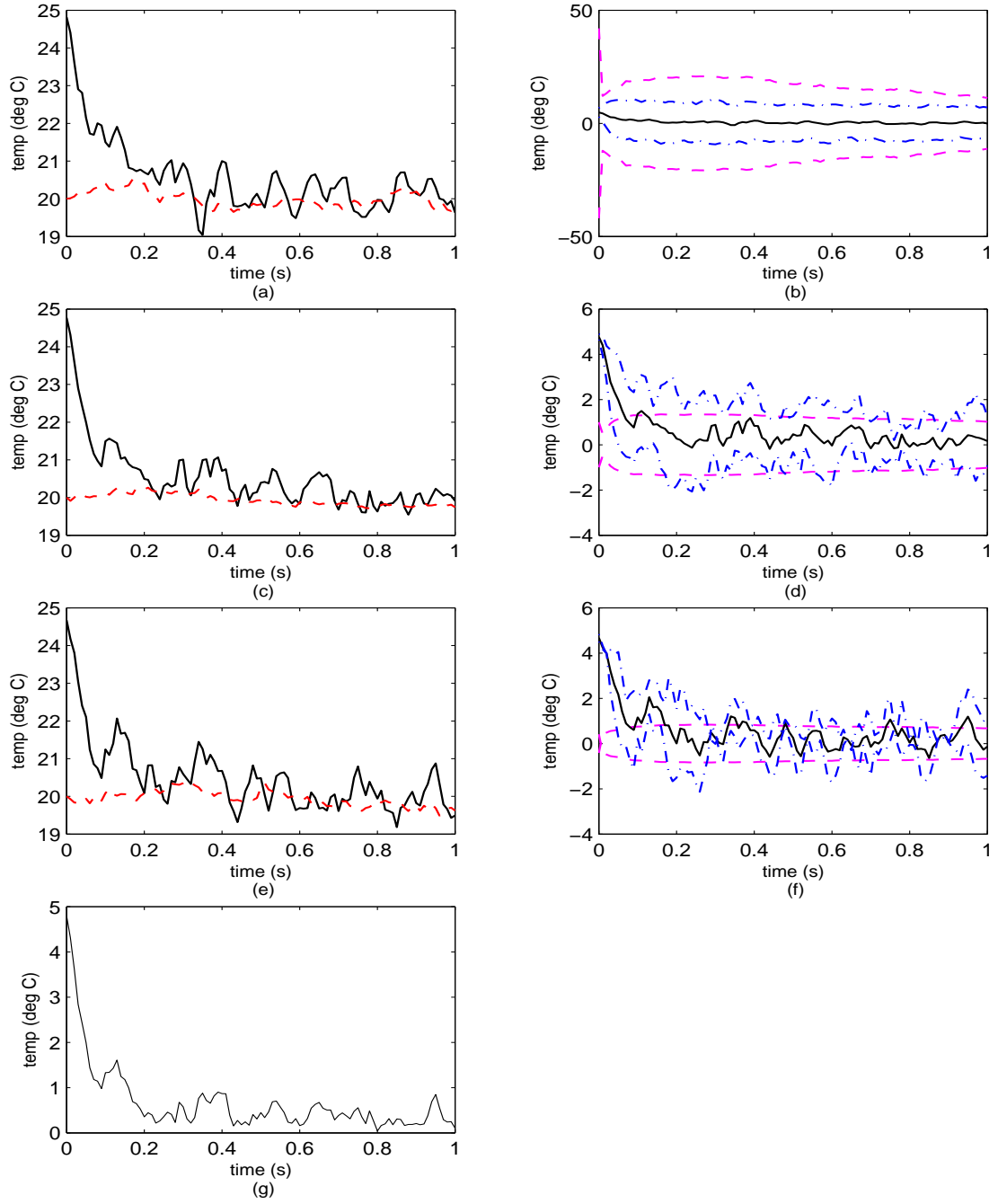


Figure 5.17 Simulation 2, Case 1 (P_0^{poor}): Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

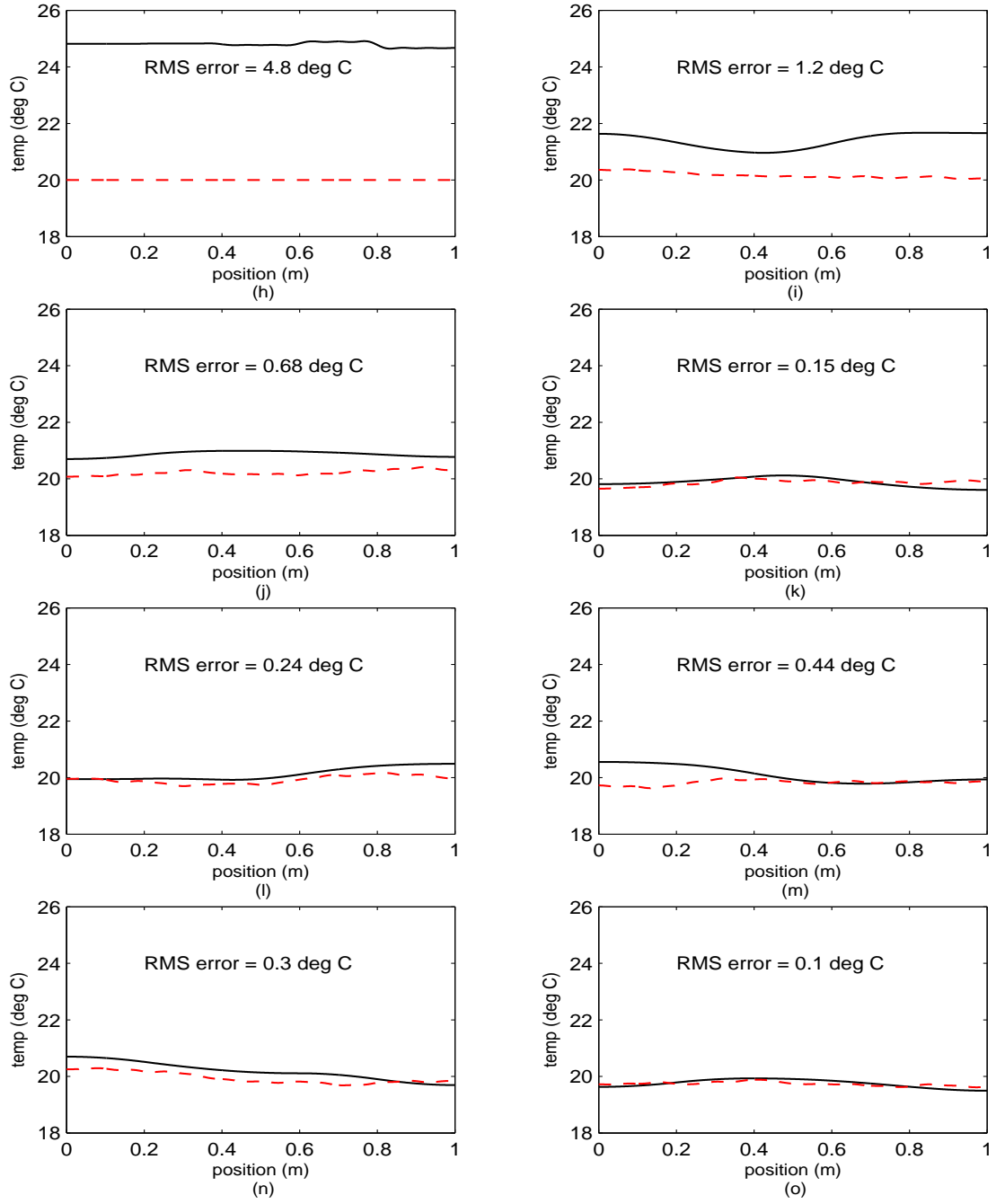


Figure 5.17 Simulation 2, Case 1 (P_0^{poor}): Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

Case 2: A good initial state covariance and a set of parameter values for the bank of elemental filters that is centered on the true dynamics noise strength. The trends from Case 1 generally hold here, except now that we have an honest appraisal of the initial state covariance, the filter can properly respond to the initial state estimate bias that it finds. We note in particular that the filters are now operating within the zero plus and minus one sigma bounds (gray dashed line) created by the initial state covariance as viewed in Figure 5.18(d, f). For proper operation, the filter must operate inside of these bounds. The results are dramatically different from those of Case 1. In Figure 5.18(a - f), we see that the state estimate converges very rapidly, even though the elemental filter (number 1) overestimates the quality of the dynamics model. In fact, the convergence is so swift, that all three elemental filters for this case compute an acceptable state estimate in just a portion of the simulated time period. The RMS error is reduced by about 90% for the first two elemental filters by the second time slice, as seen in Figures 5.18(j) and 5.20(j). As in Case 1, elemental filter 3, [Figure 5.20(a - f)] converges very rapidly and then overshoots and continually overreacts to the noise entering through the measurements — the initial covariance estimate appears to have little bearing on a filter based on a model with “high” Q . A high assumed value for Q “flushes” the initial conditions out of the system because the gain is high when Q is high, i.e., low confidence in the dynamics model or simply a large amount of process noise.

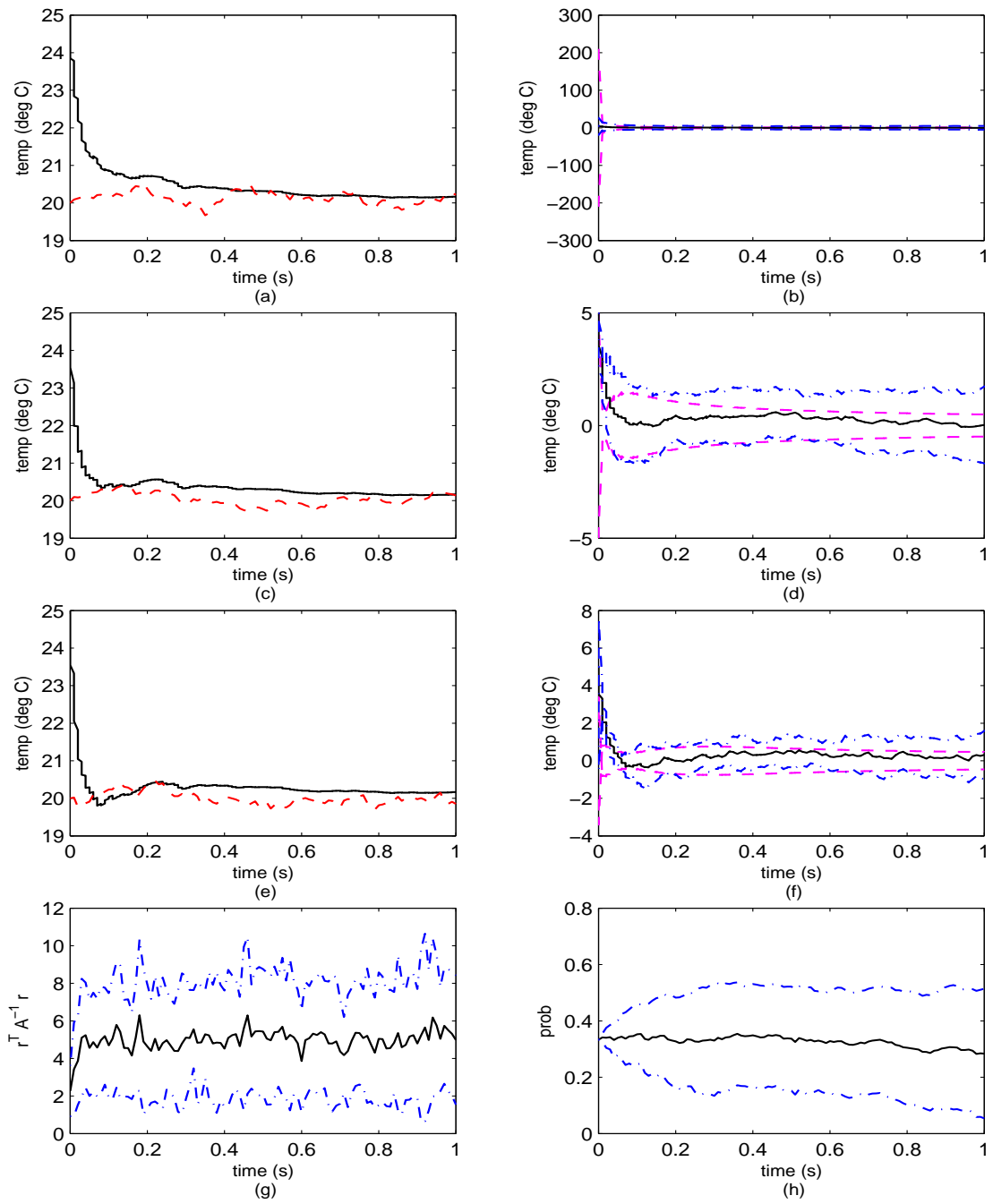


Figure 5.18 Simulation 2, Case 2 (P_0^{good}): Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

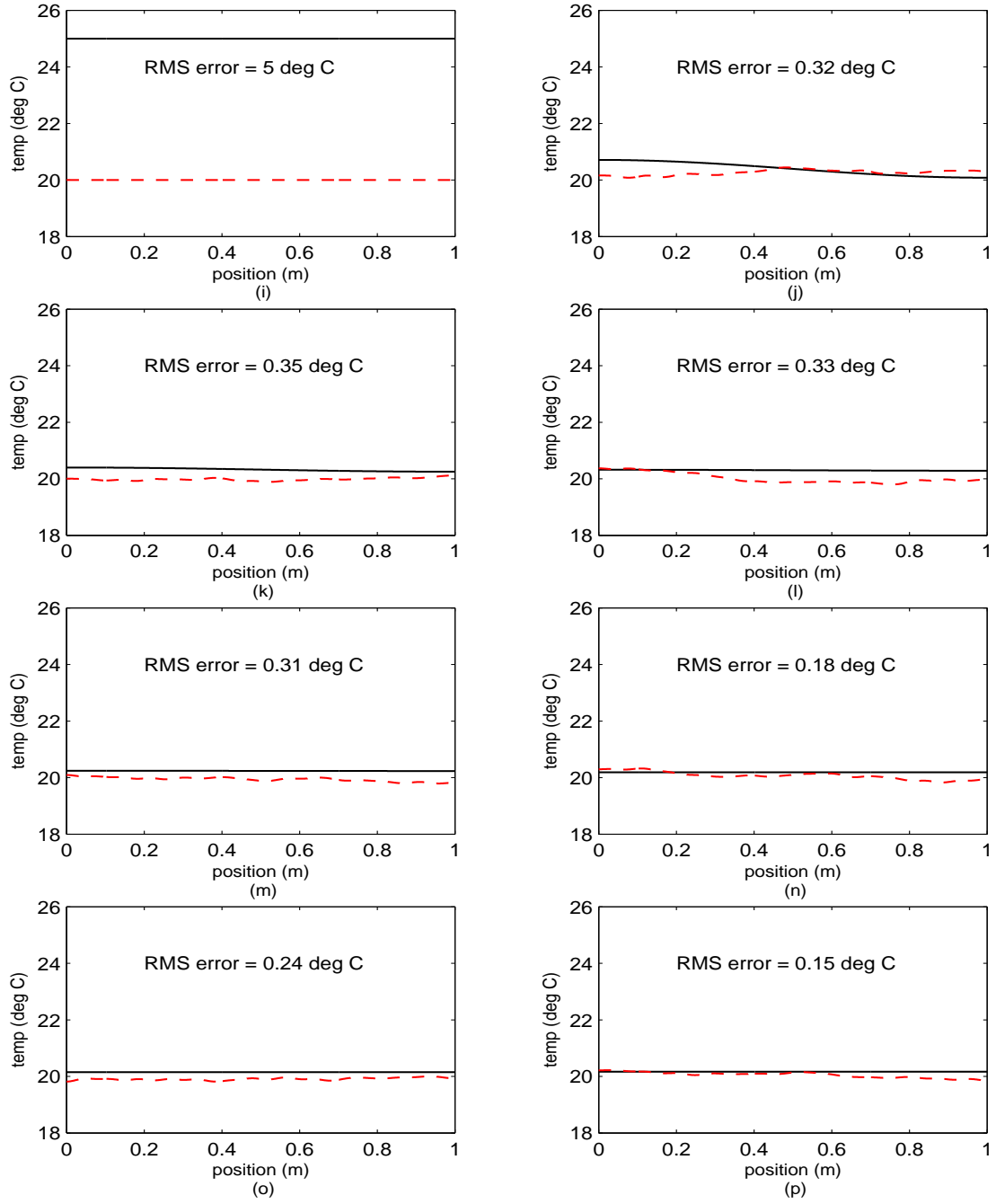


Figure 5.18 Simulation 2, Case 2 (P_0^{good}): Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

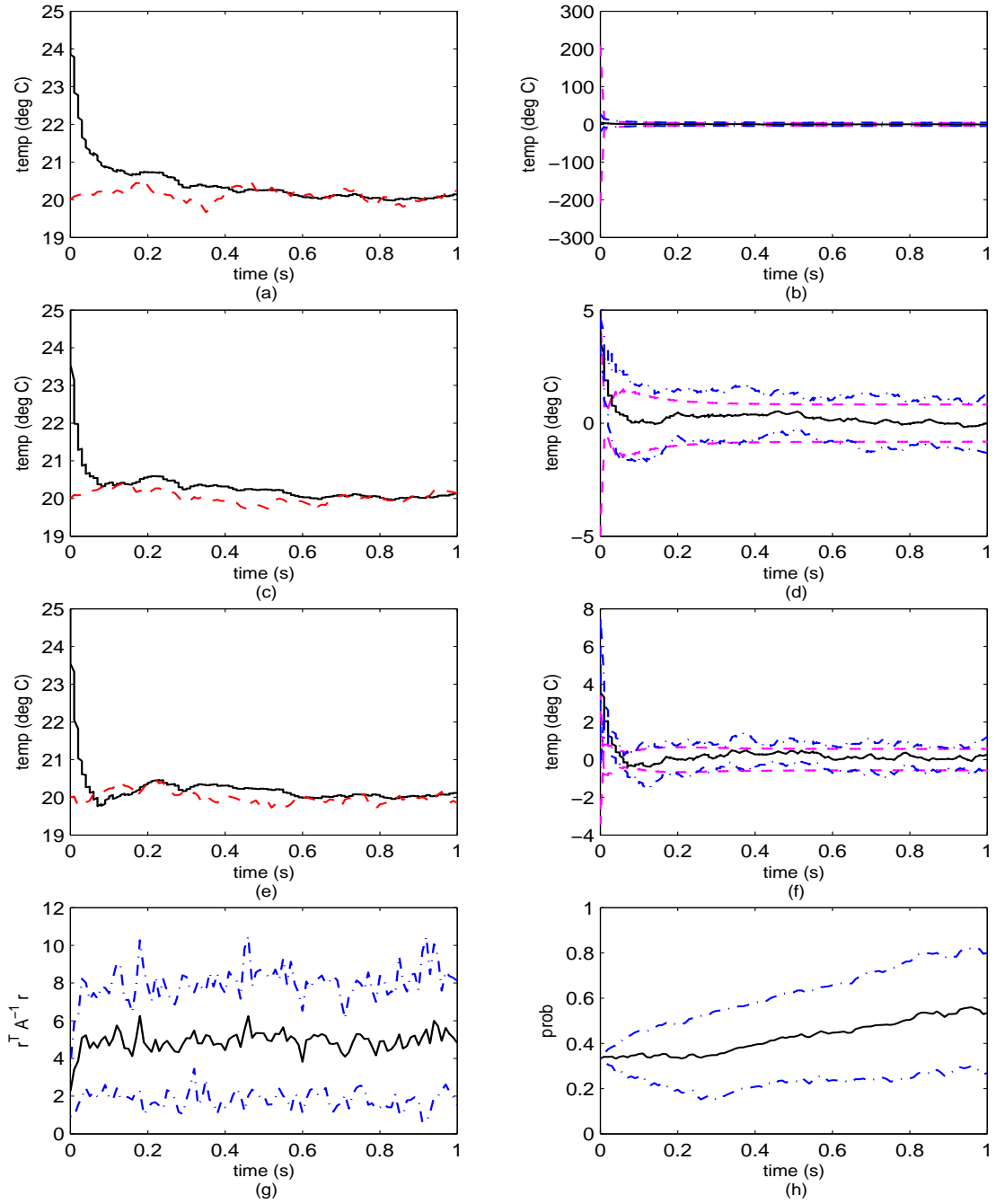


Figure 5.19 Simulation 2, Case 2 (P_0^{good}): Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

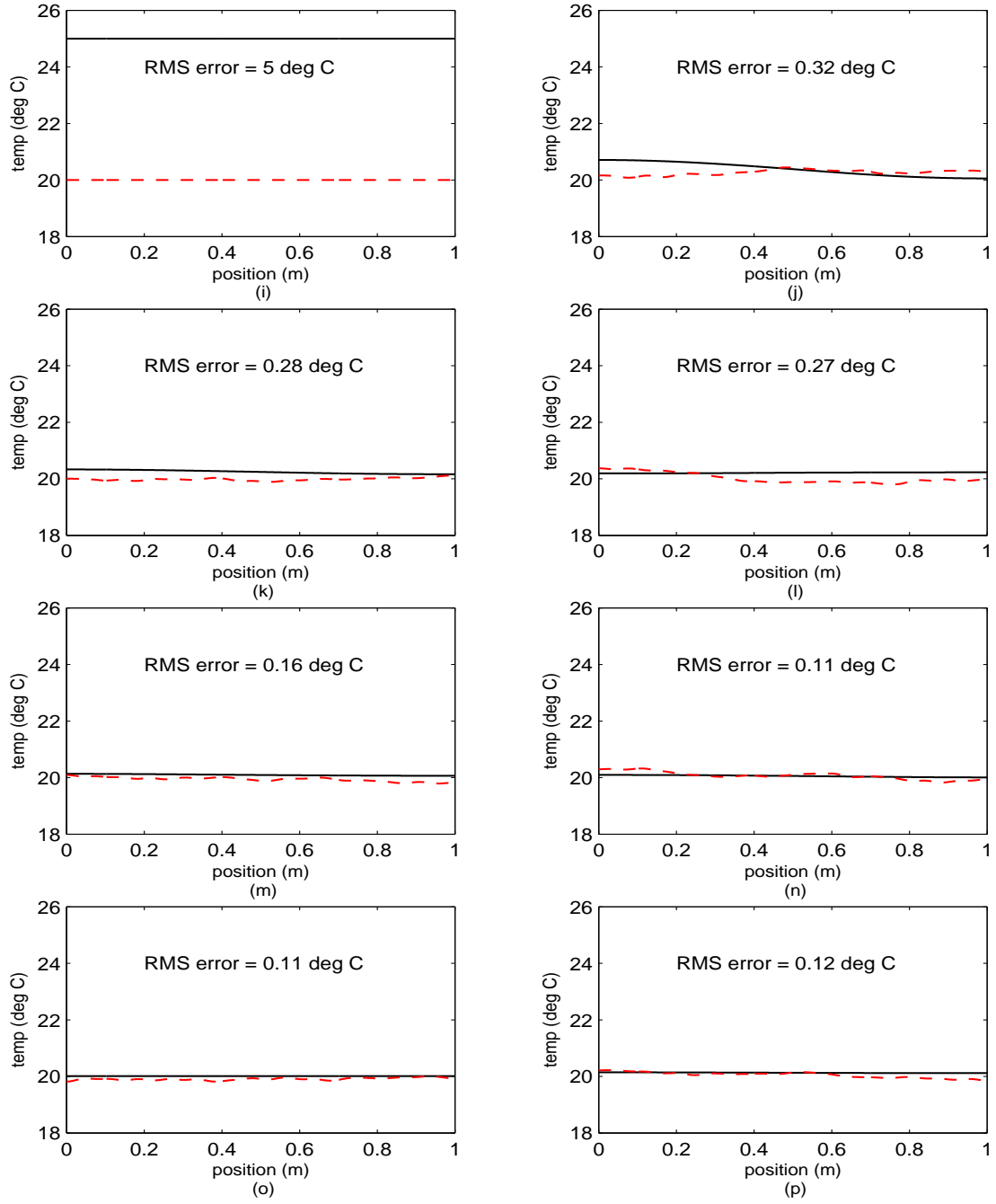


Figure 5.19 Simulation 2, Case 2 (P_0^{good}): Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

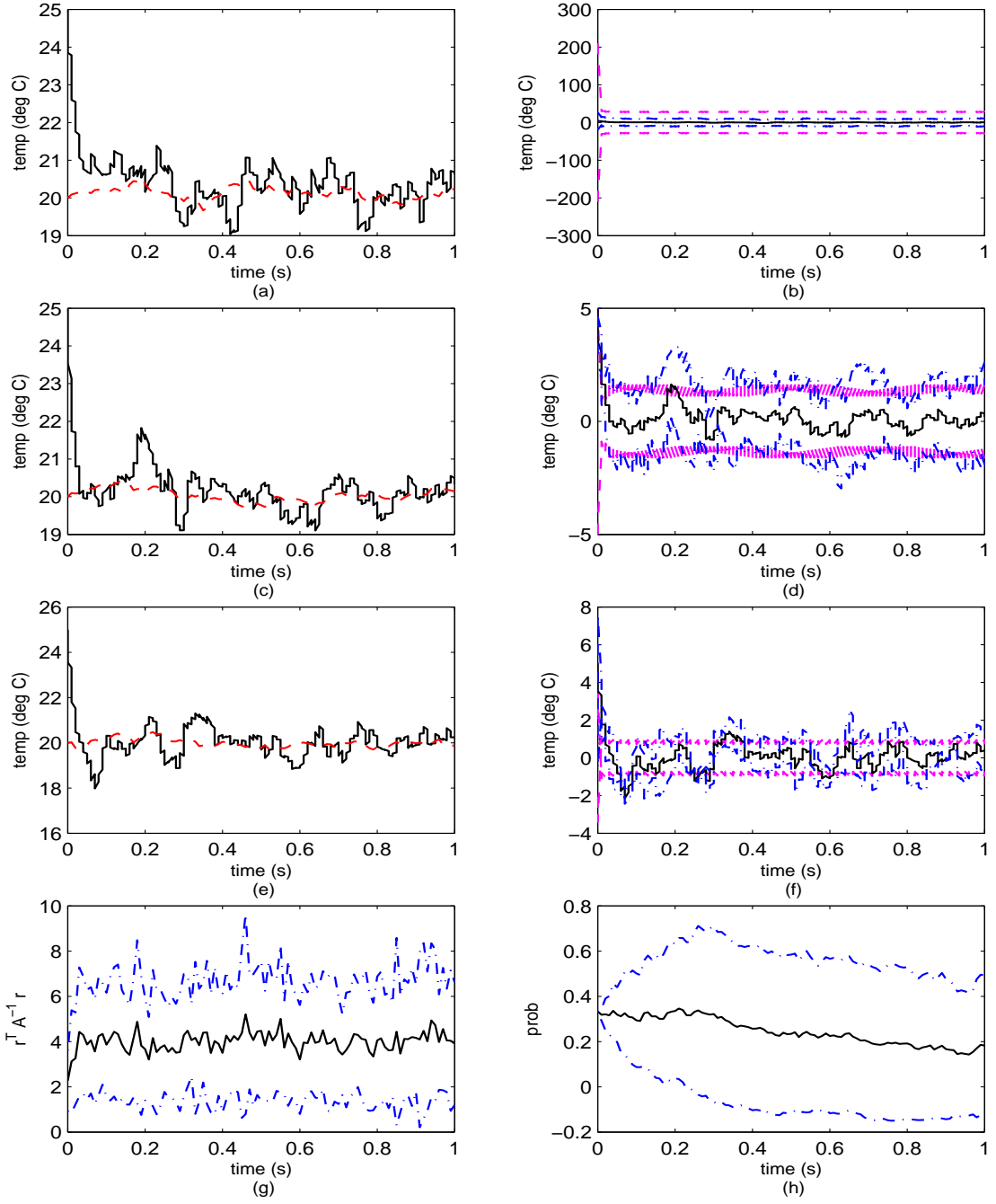


Figure 5.20 Simulation 2, Case 2 (P_0^{good}): Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

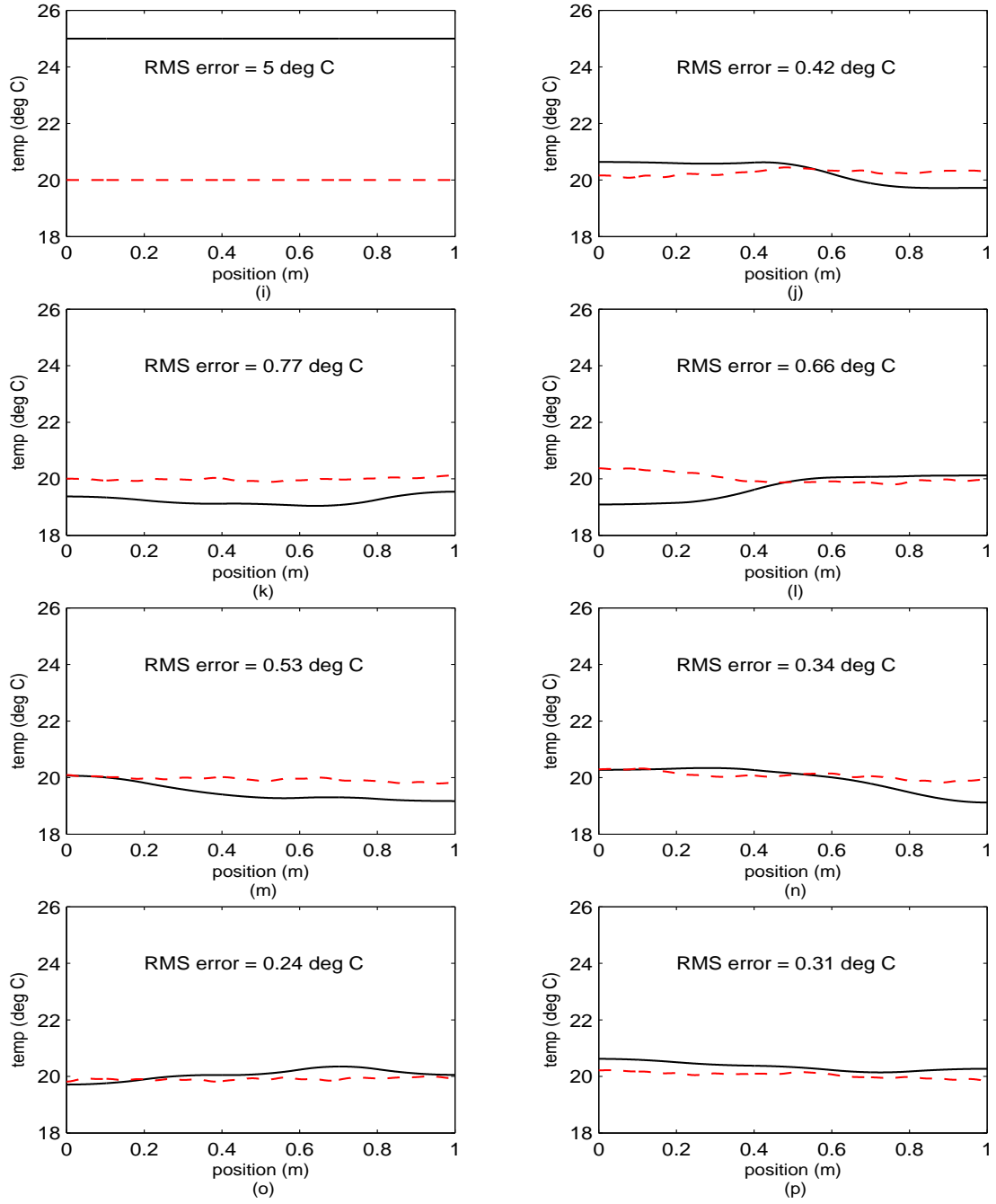


Figure 5.20 Simulation 2, Case 2 (P_0^{good}): Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

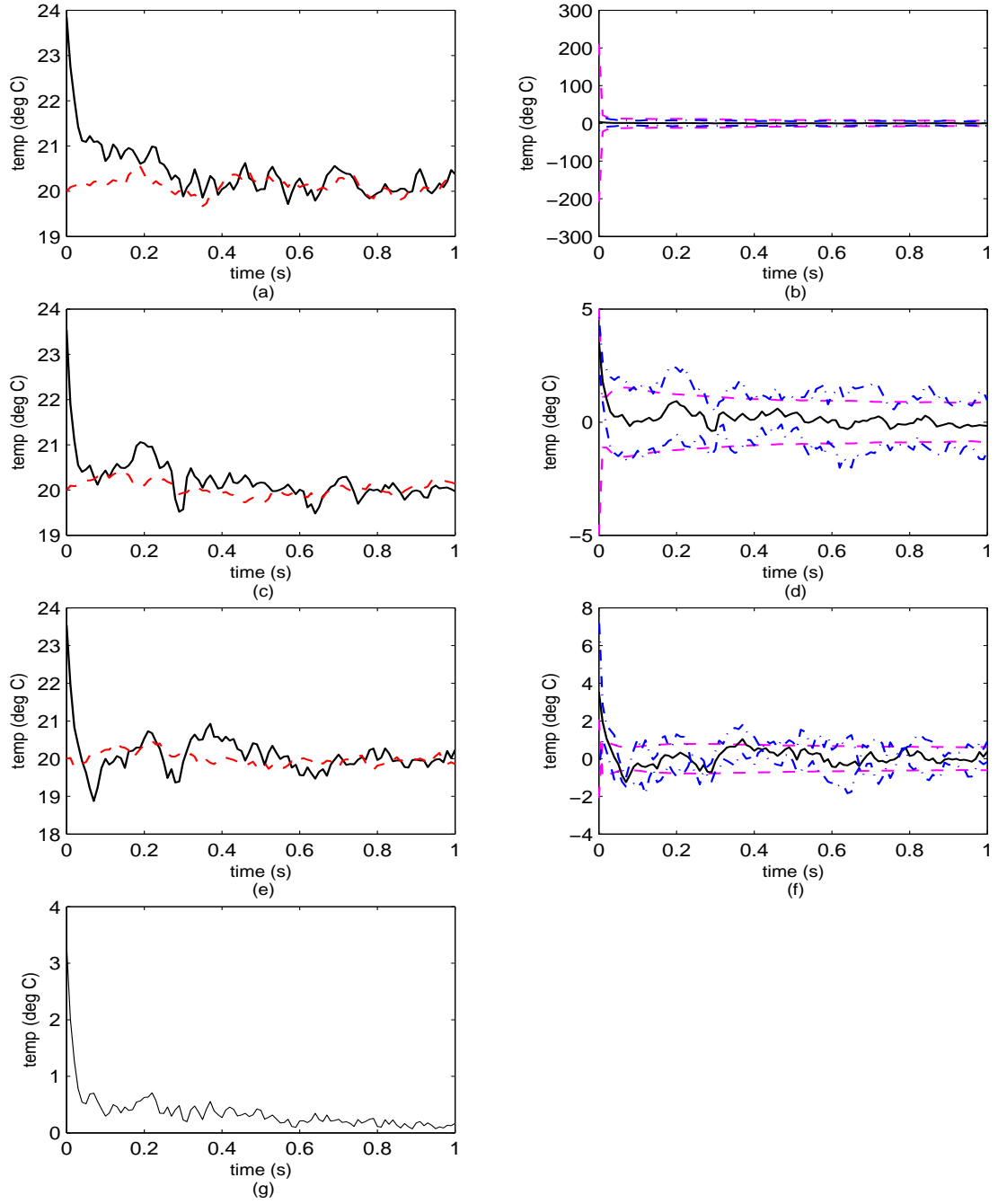


Figure 5.21 Simulation 2, Case 2 (P_0^{good}): Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

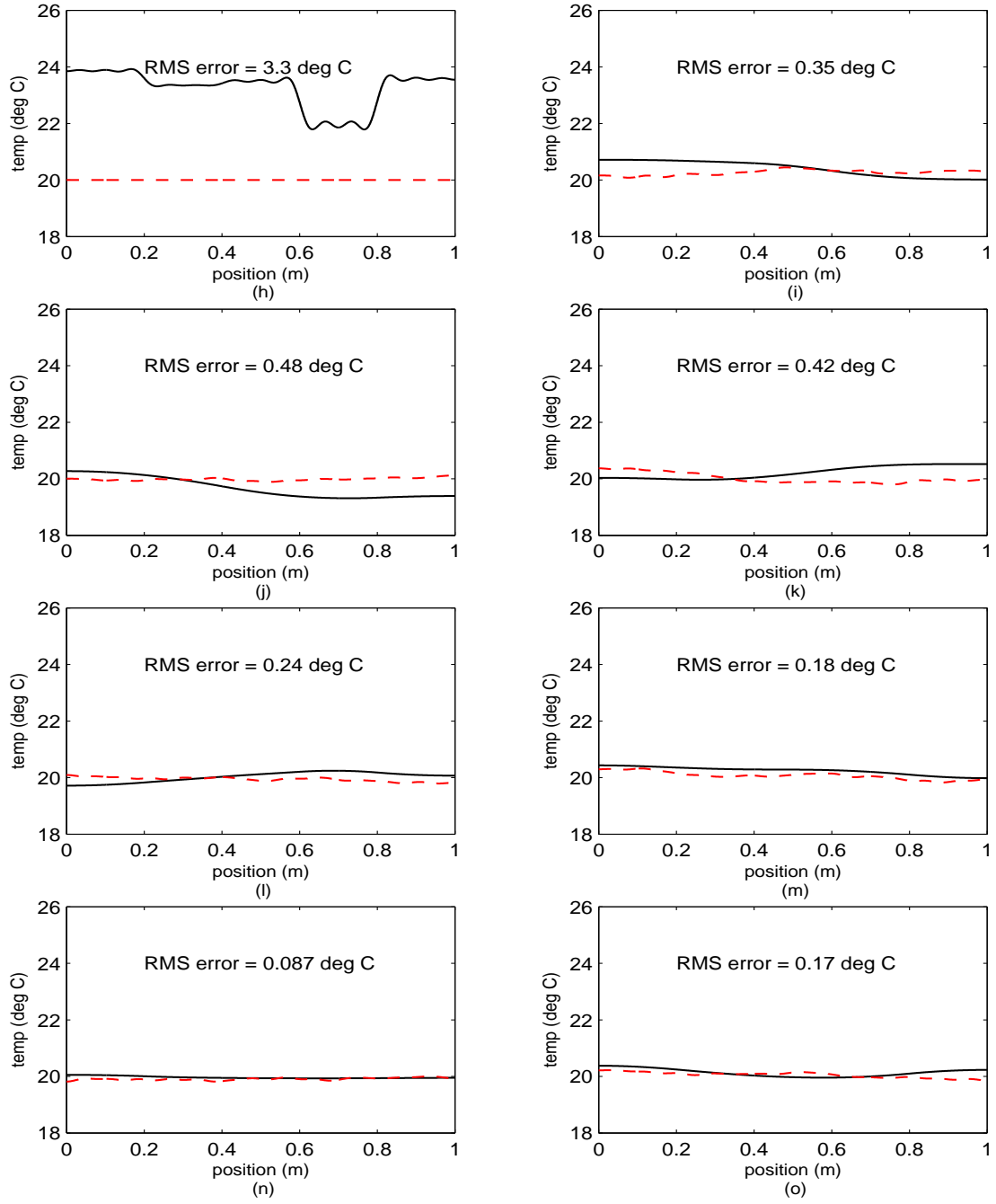


Figure 5.21 Simulation 2, Case 2 (P_0^{good}): Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

Case 3: A poor initial state covariance and a set of parameter values for the bank of elemental filters that is centered at ten times the true dynamics noise strength.

A close inspection of Figure 5.22(a - f) for elemental filter 1 shows that the state estimate is slowly converging to the true state just as it did in Case 1, Figure 5.14. Again, we observe at time zero in Figure 5.22(d, f) that the state estimate mean error (solid line) and mean plus and minus one sigma (dash-dot line) are outside the zero plus and minus one filter-computed sigma bounds (gray dashed line) created by the initial state covariance. The estimate progresses along the entire rod as shown in Figure 5.22(i - p). By comparison, for elemental filter 2 in Figure 5.23(a - f), we see that the state estimate converges more quickly than for elemental filter 2 in Figure 5.15(a - f) because we have a larger assumed value for the dynamics noise strength. As we have previously noted, a larger filter-assumed Q , for a given R , gives rise to a larger gain and is thus more responsive to the measurements. The model for elemental filter 3 greatly overstates the true dynamics noise strength, see Figure 5.24(a - f)], and it converges very rapidly and then overshoots and continually overreacts to the noise entering through the measurements; as before, the initial covariance estimate appears to have little bearing on the performance of a filter based on this model, compared to the ill match provided by overstating Q_{true} . When the filter's Q is much too high, the transient caused by the initial conditions is very short and then filter performance relies heavily on the measurements. All of these attributes contribute to driving the elemental filter 2 mean hypothesis conditional probability towards one.

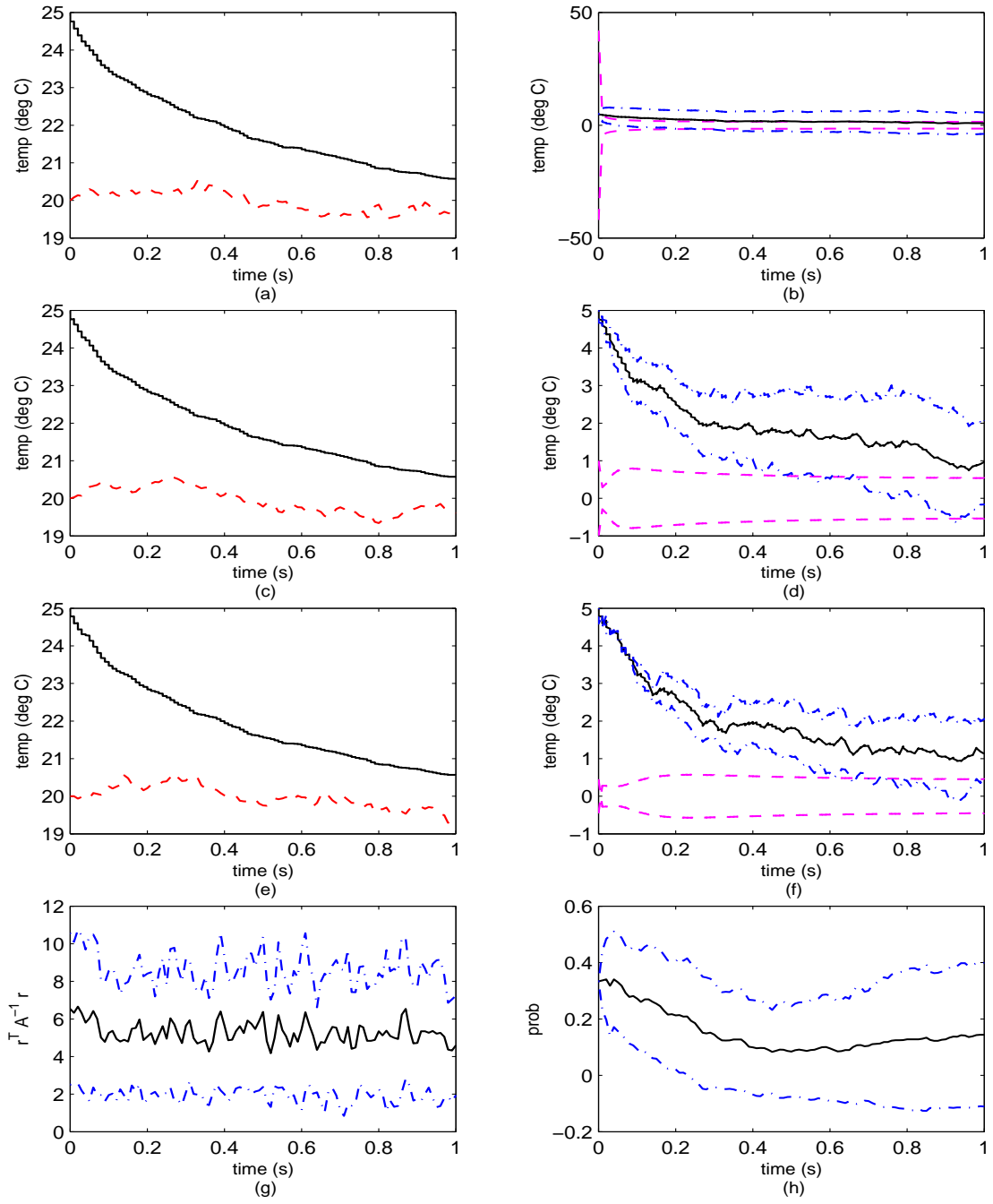


Figure 5.22 Simulation 2, Case 3 (P_0^{poor}): Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

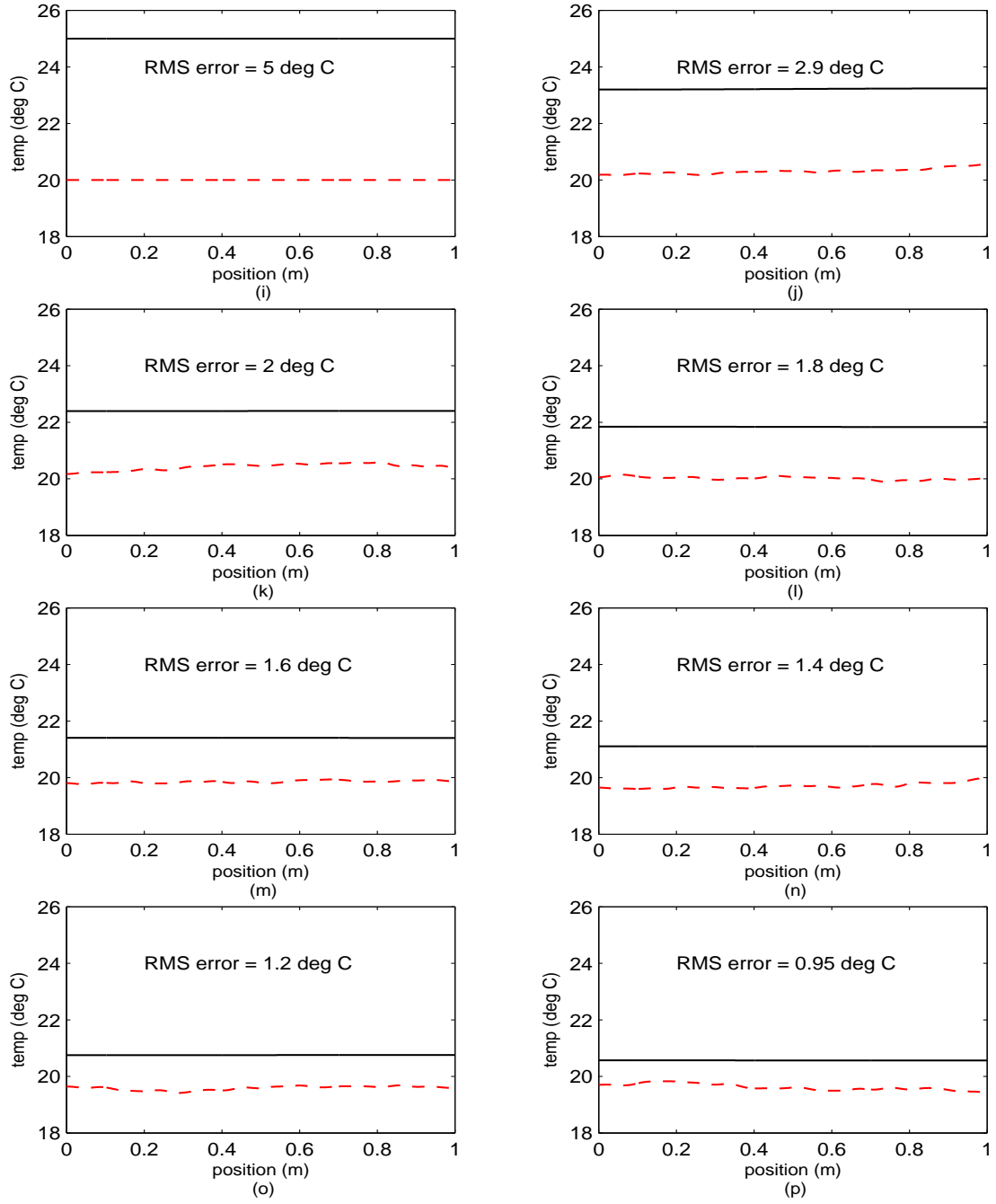


Figure 5.22 Simulation 2, Case 3 (P_0^{poor}): Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

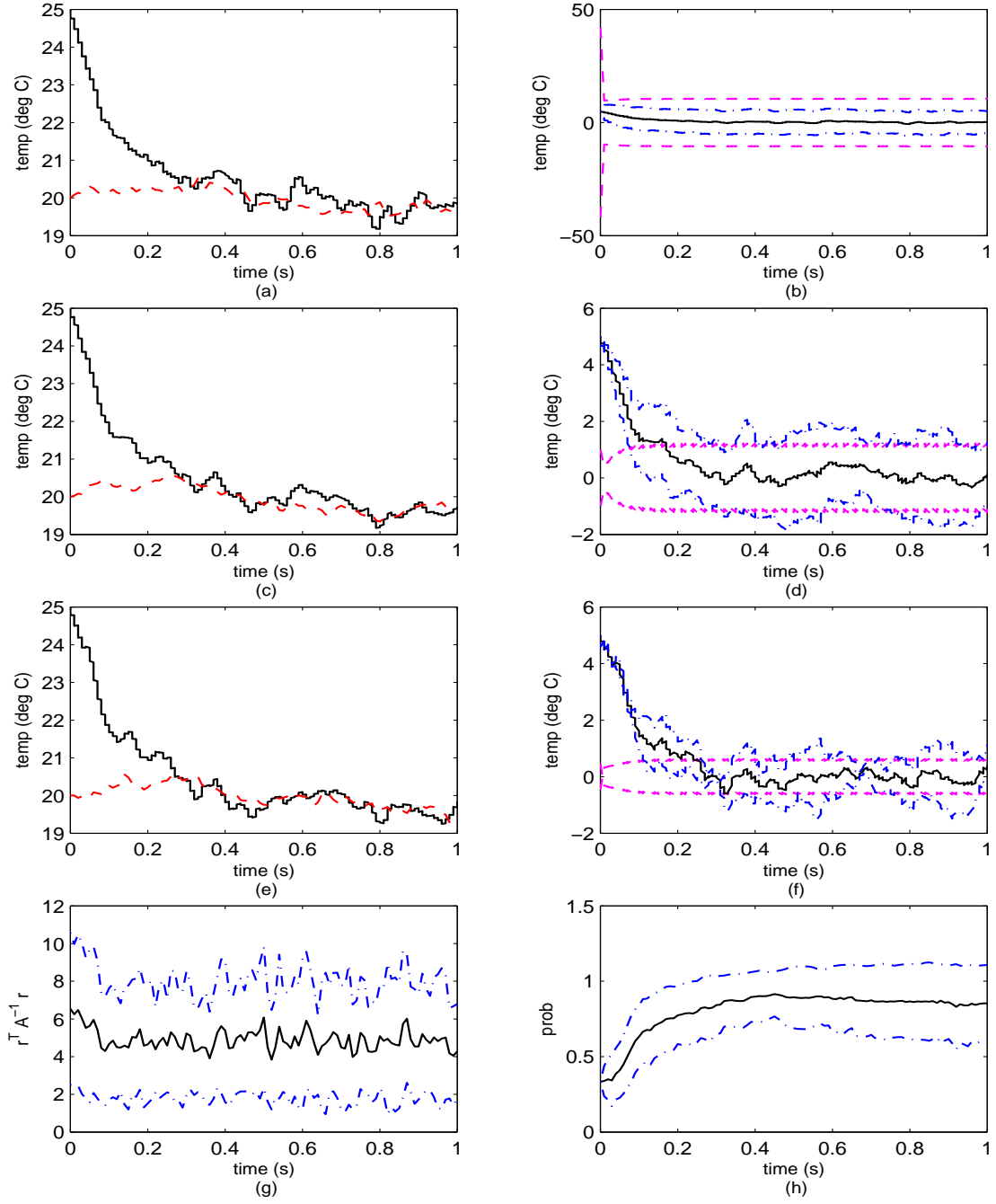


Figure 5.23 Simulation 2, Case 3 (P_0^{poor}): Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

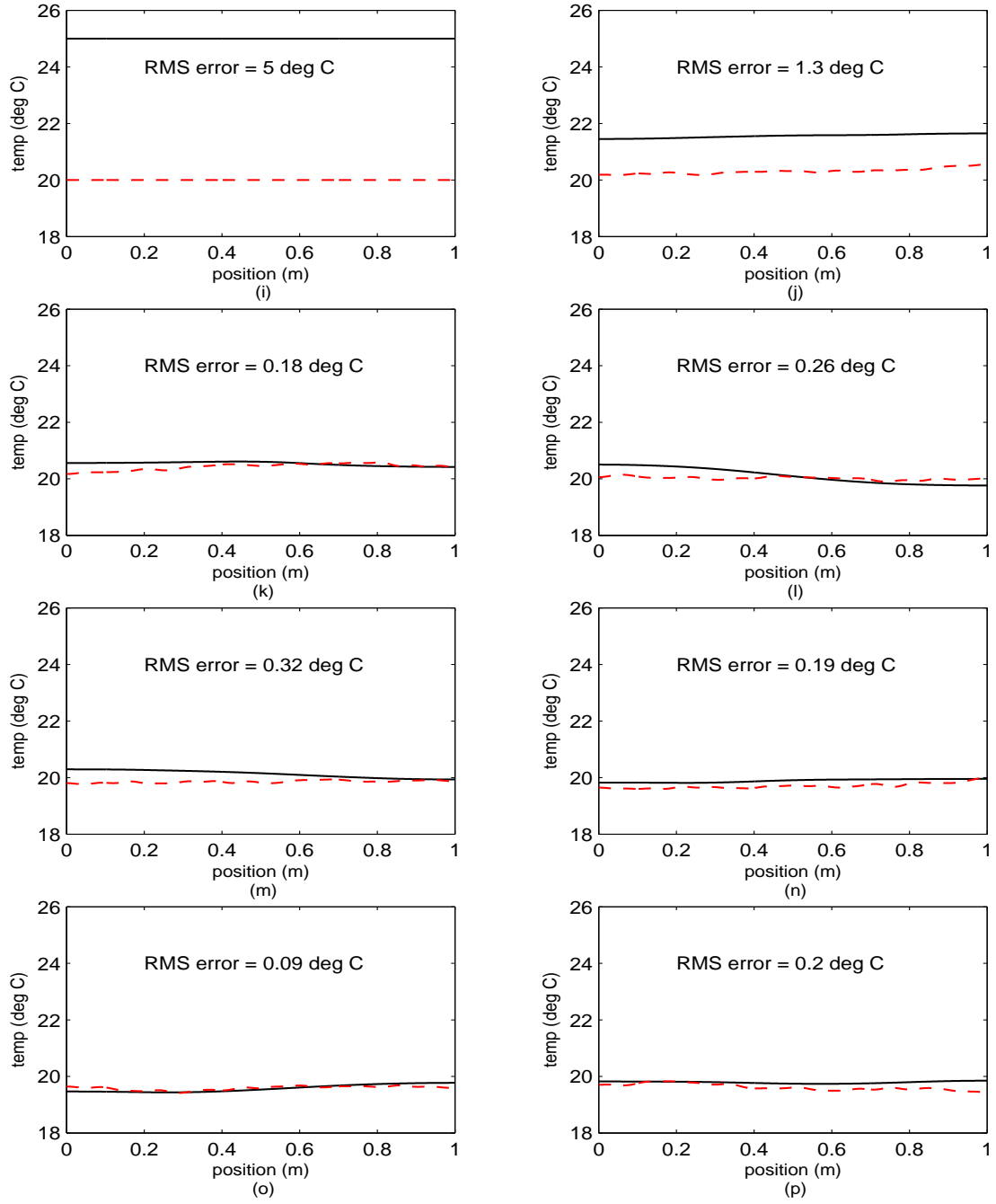


Figure 5.23 Simulation 2, Case 3 (P_0^{poor}): Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

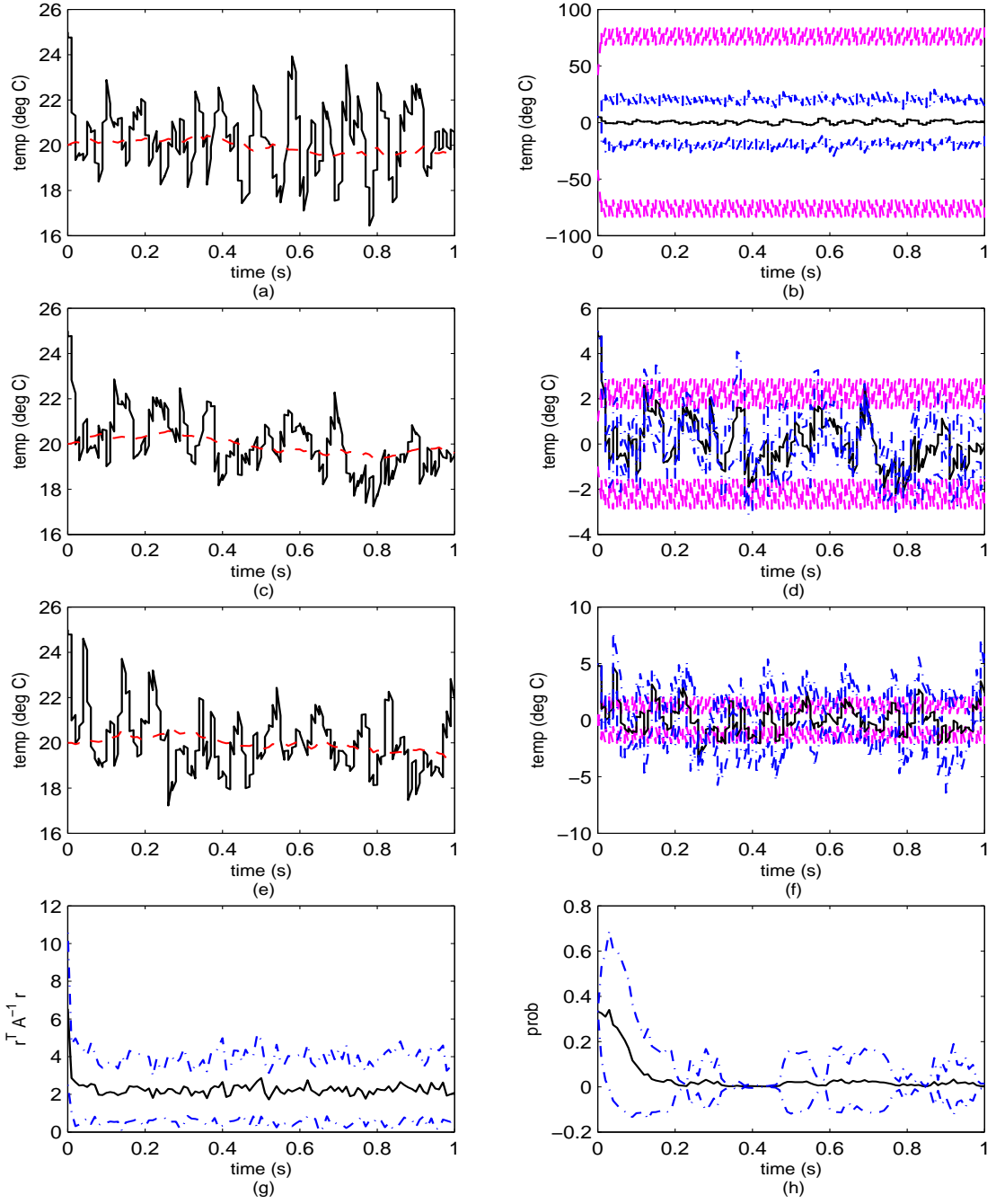


Figure 5.24 Simulation 2, Case 3 (P_0^{poor}): Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

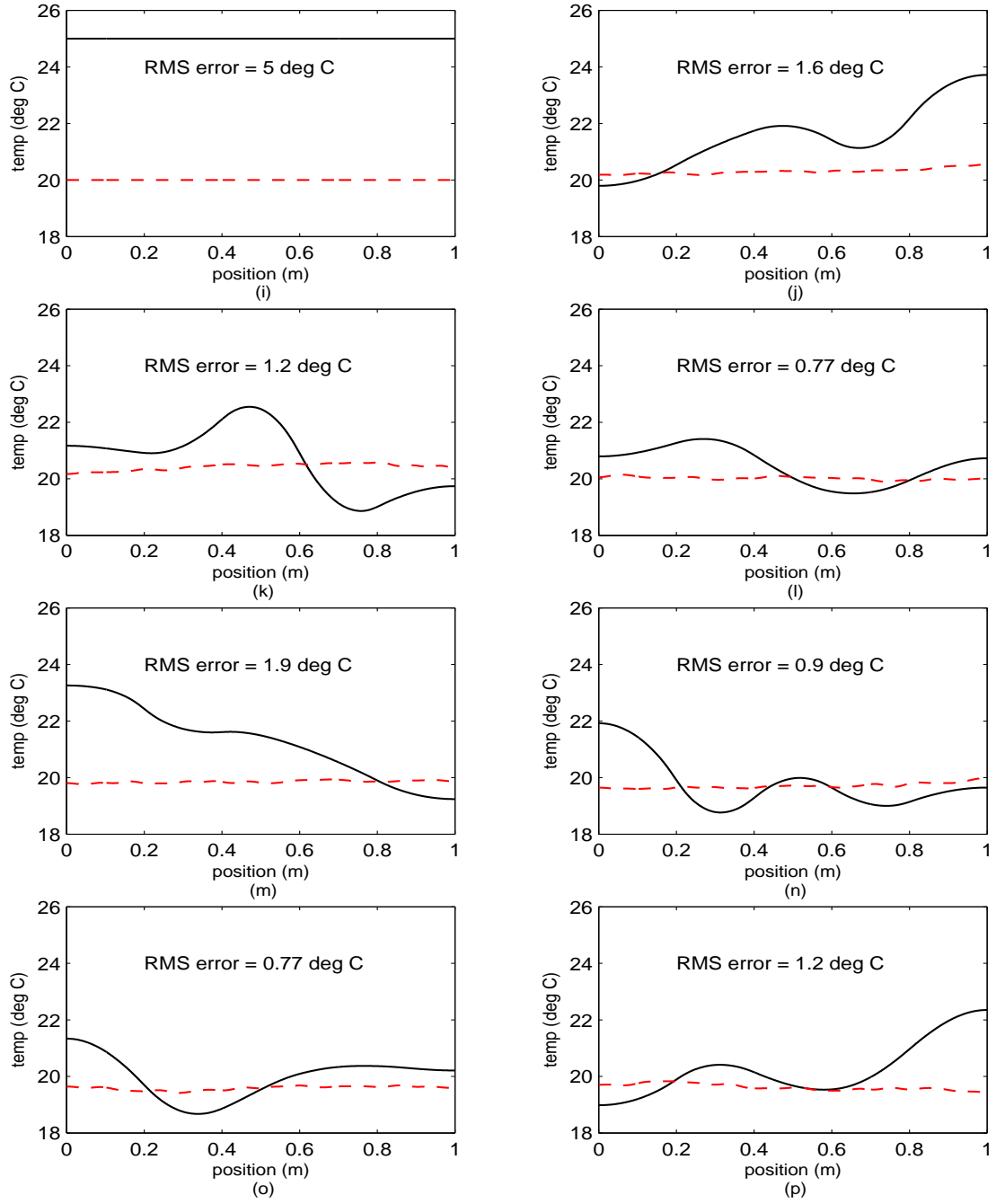


Figure 5.24 Simulation 2, Case 3 (P_0^{poor}): Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

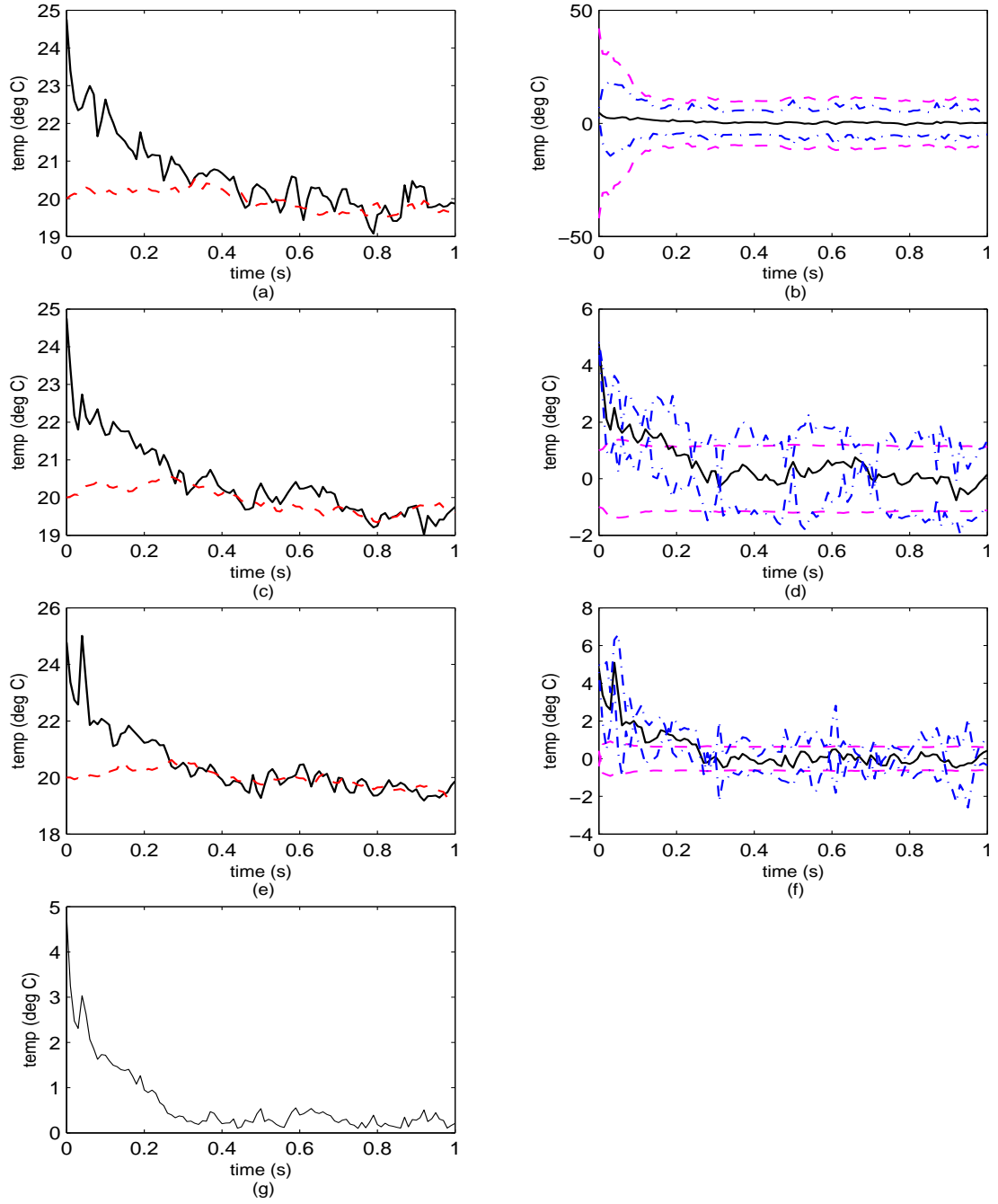


Figure 5.25 Simulation 2, Case 3 (P_0^{poor}): Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

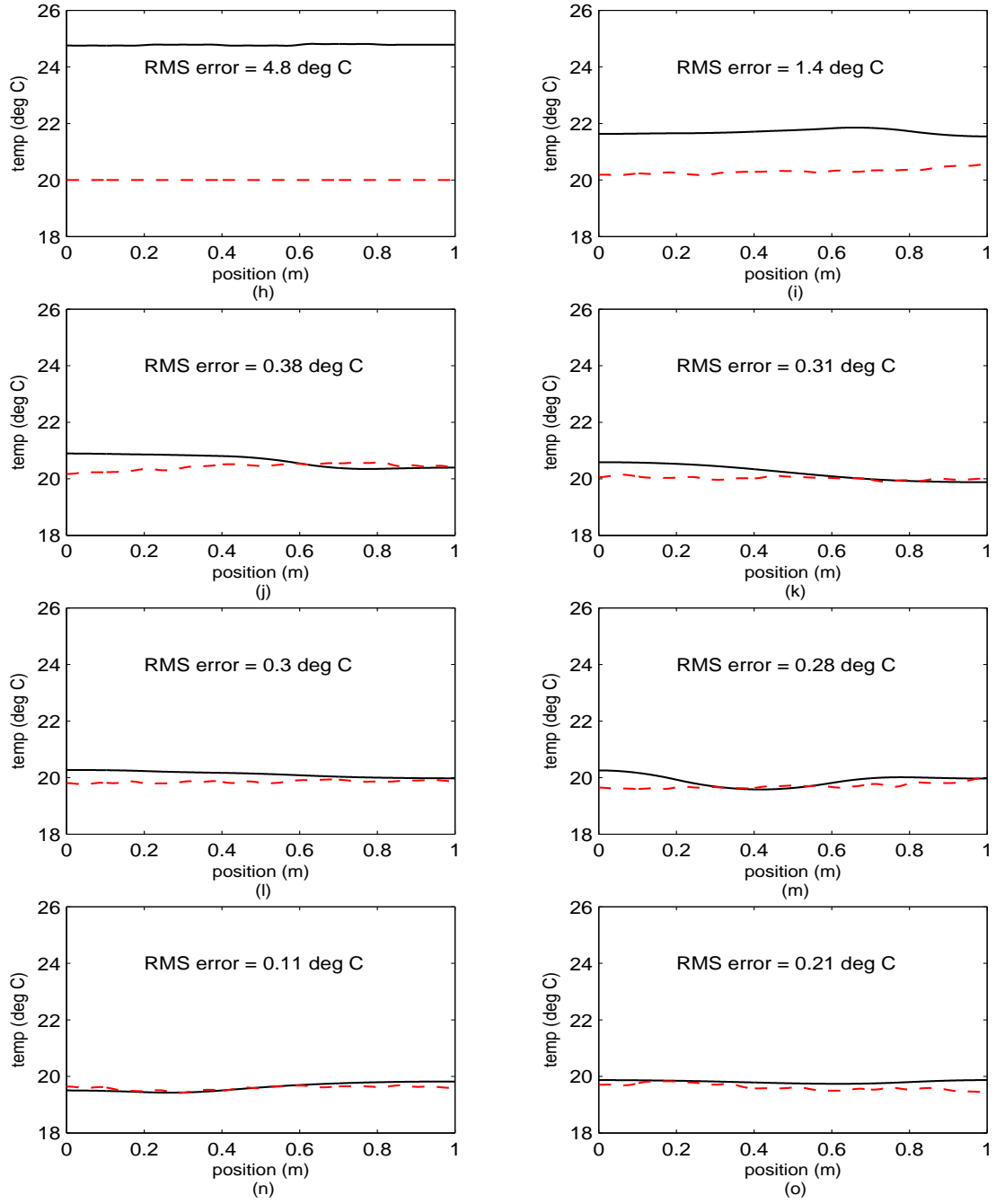


Figure 5.25 Simulation 2, Case 3 (P_0^{poor}): Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

Case 4: A good initial state covariance and a set of parameter values for the bank of elemental filters that is centered at ten times the true dynamics noise strength. Now that we have a good initial state covariance, elemental filter 1 appears to be a better match relative to the second elemental filter because its filter-assumed low (optimistic) assessment of the dynamics noise strength is tempered by a good assumed value for the initial state covariance as contrasted by Case 3 where we had a poor choice for P_0 . The poor (too small) choice for P_0 in Case 3 made it easier for the MMAE to recognize the first elemental filter with a too-low value for Q (and thus slow dynamic response for the corresponding elemental filter) was based on the wrong model; hence the low hypothesis conditional probability. To see the interplay of the initial state covariance and the assumed dynamics noise strength more clearly, we turn to Equation (5.11), which shows us that the updated state covariance, seeded by P_0 , is “scaled” by the (“square” of the) state transition matrix and then added to the discrete-time version of the dynamics noise strength, \mathbf{Q}_d to accomplish time propagation. Thus, the strength of the noise for the first filter appears higher and thus it often accounts for the variance in the system and thus is an attractive choice for the MMAE. However, state estimation performance would be degraded if the initial state covariance were purposefully set lower than the true value.

When an elemental filter is based on an exceedingly high assumed value for Q (e.g., *three* orders of magnitude greater than Q_{true}), the MMAE quickly recognizes the mismodeled filter. This behavior can be seen readily by comparing Cases 3 and 4 to Cases 1 and 2 in which the third elemental filter is more conservatively set at *just* two orders of magnitude greater than the true value. Since the measurement covariance, R , is assumed known and is used to create all of the filters in the bank, the corresponding gain in these filters is relatively high and thus the filters attempt to follow the measurements. This situation creates a mismatch between the filter-assumed error covariance and the covariance of the residuals, as can be observed in the likelihood quotient histories for the third elemental filter in all four cases (but

more substantially for Cases 3 and 4) in Figures 5.16(g), 5.20(g), 5.24(g), and 5.28(g). When a poor choice is made for P_0 , we have seen that the MMAE can more quickly flow probability away from elemental filters with underestimated Q . Conversely, a poor choice for P_0 compounds the difficulty of identifying an elemental filter based on an assumed Q that is too high — compare the elemental filter 3 probability histories for Cases 1 and 3 versus Cases 2 and 4.

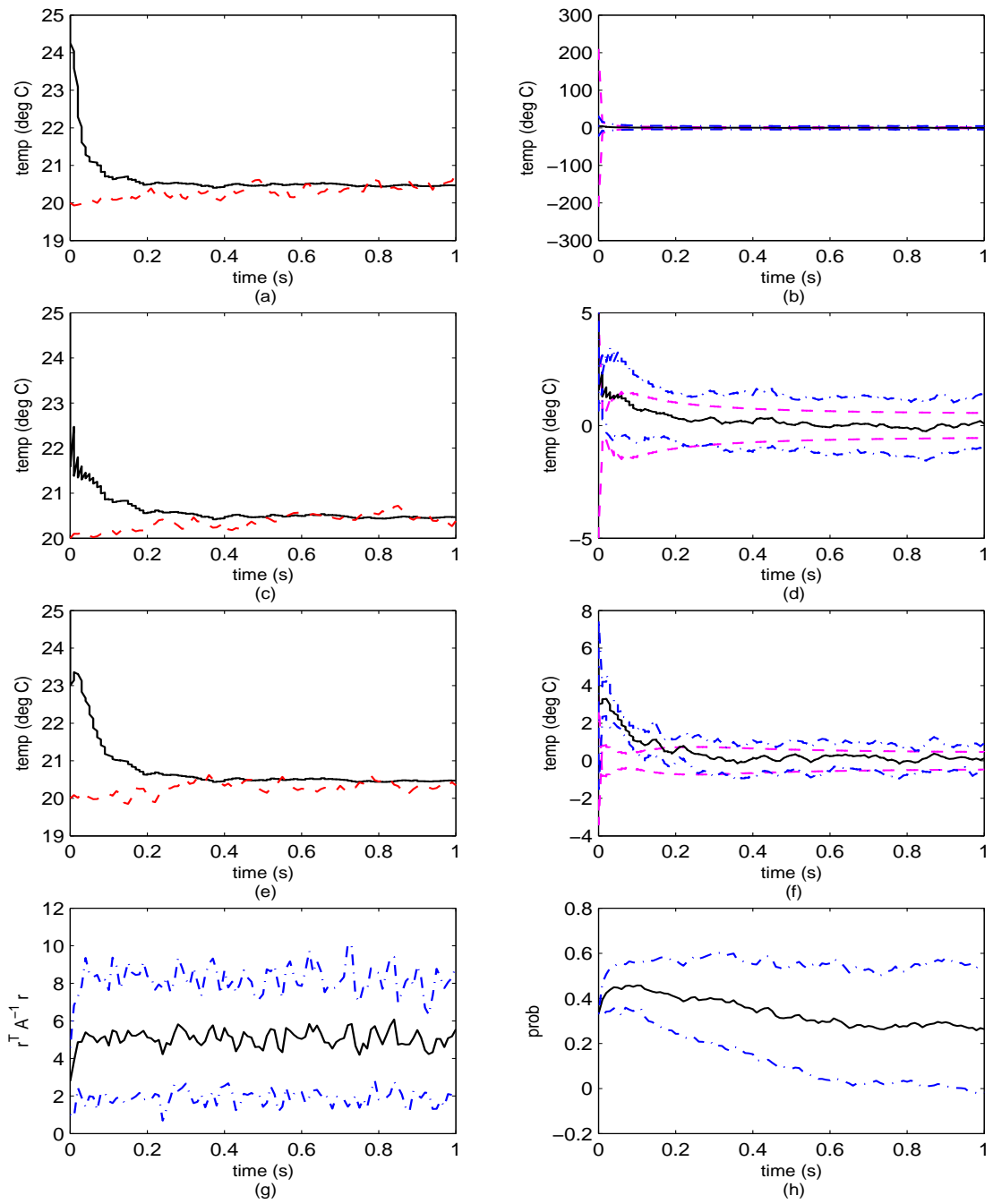


Figure 5.26 Simulation 2, Case 4 (P_0^{good}): Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

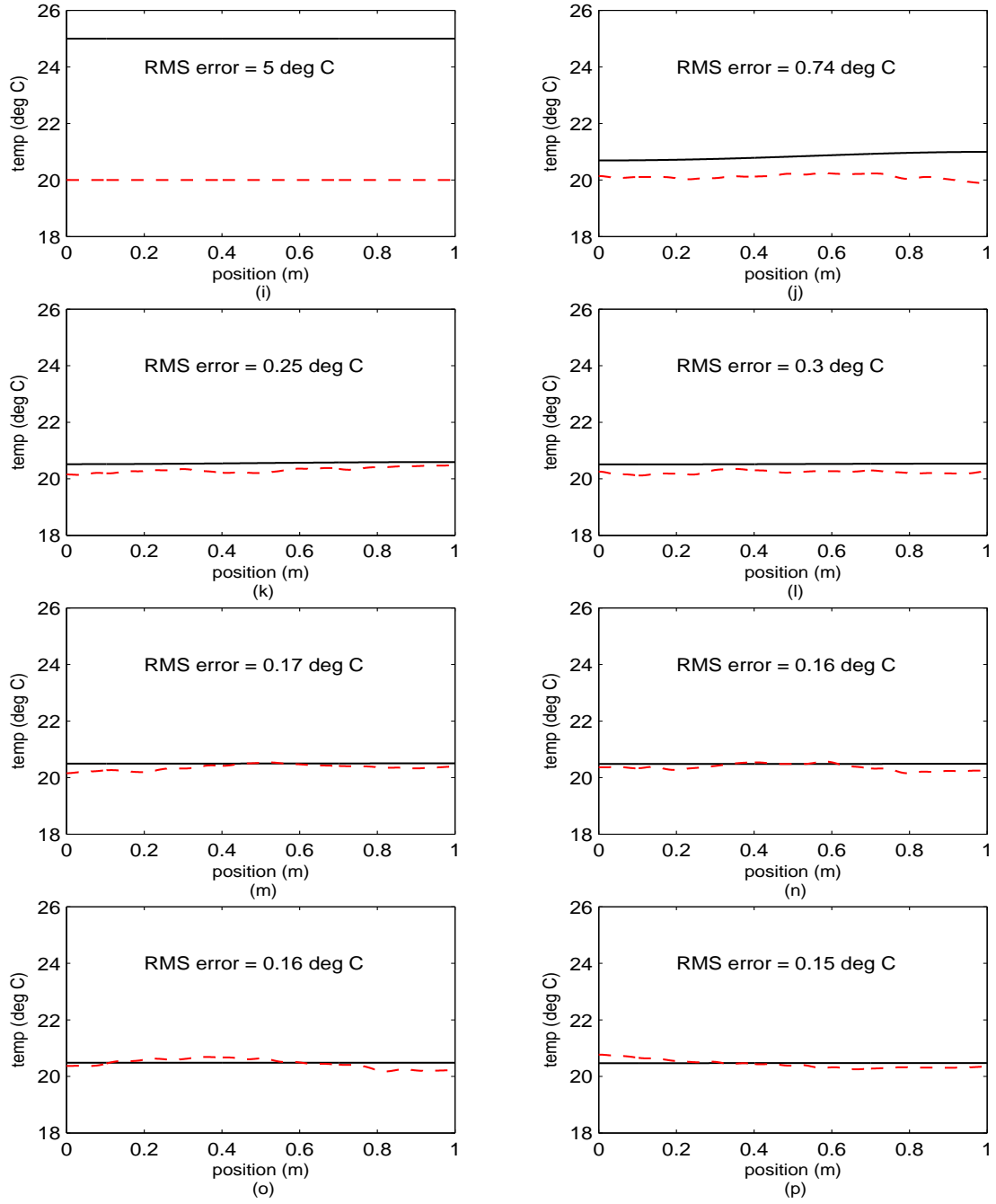


Figure 5.26 Simulation 2, Case 4 (P_0^{good}): Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

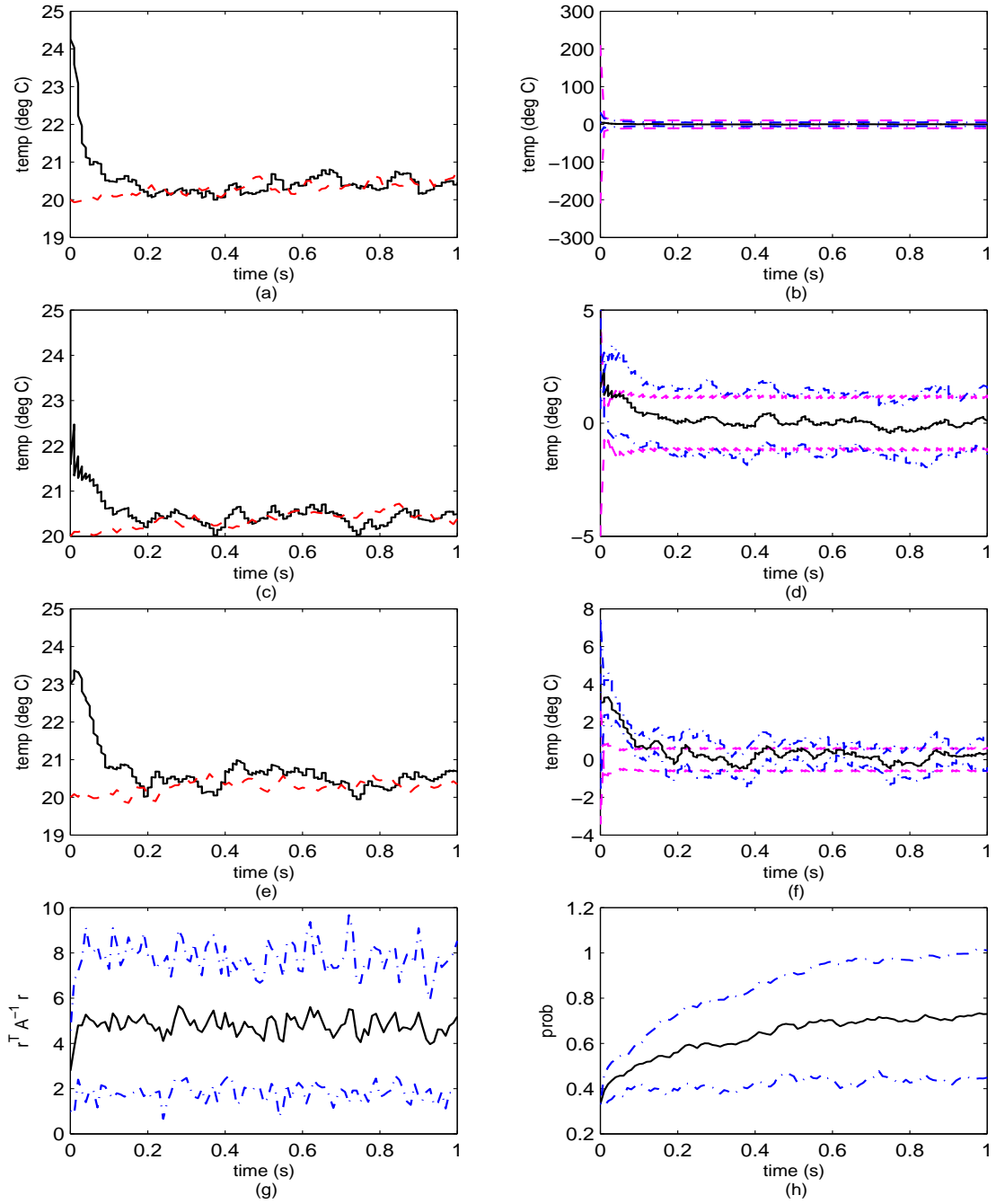


Figure 5.27 Simulation 2, Case 4 (P_0^{good}): Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

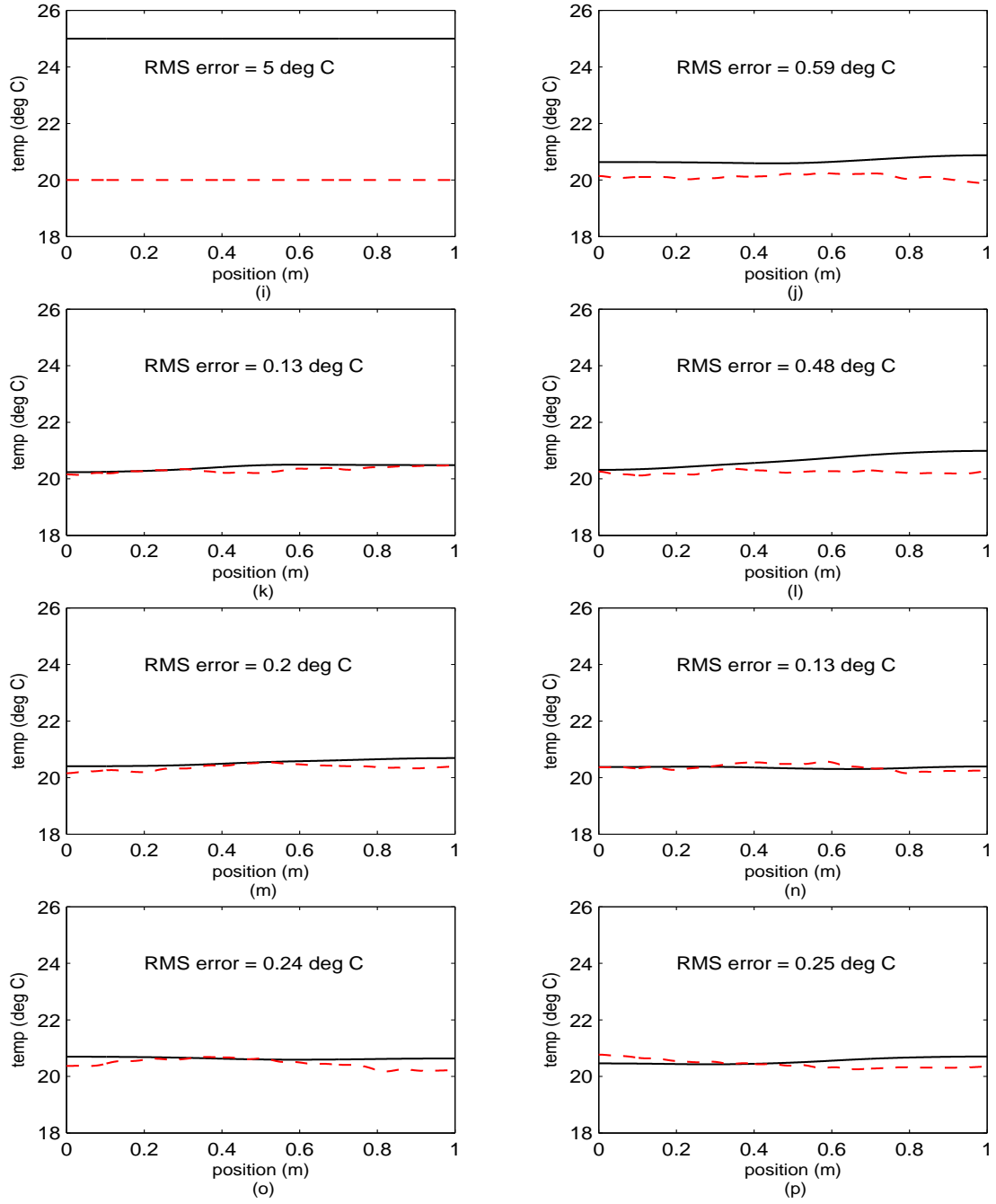


Figure 5.27 Simulation 2, Case 4 (P_0^{good}): Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

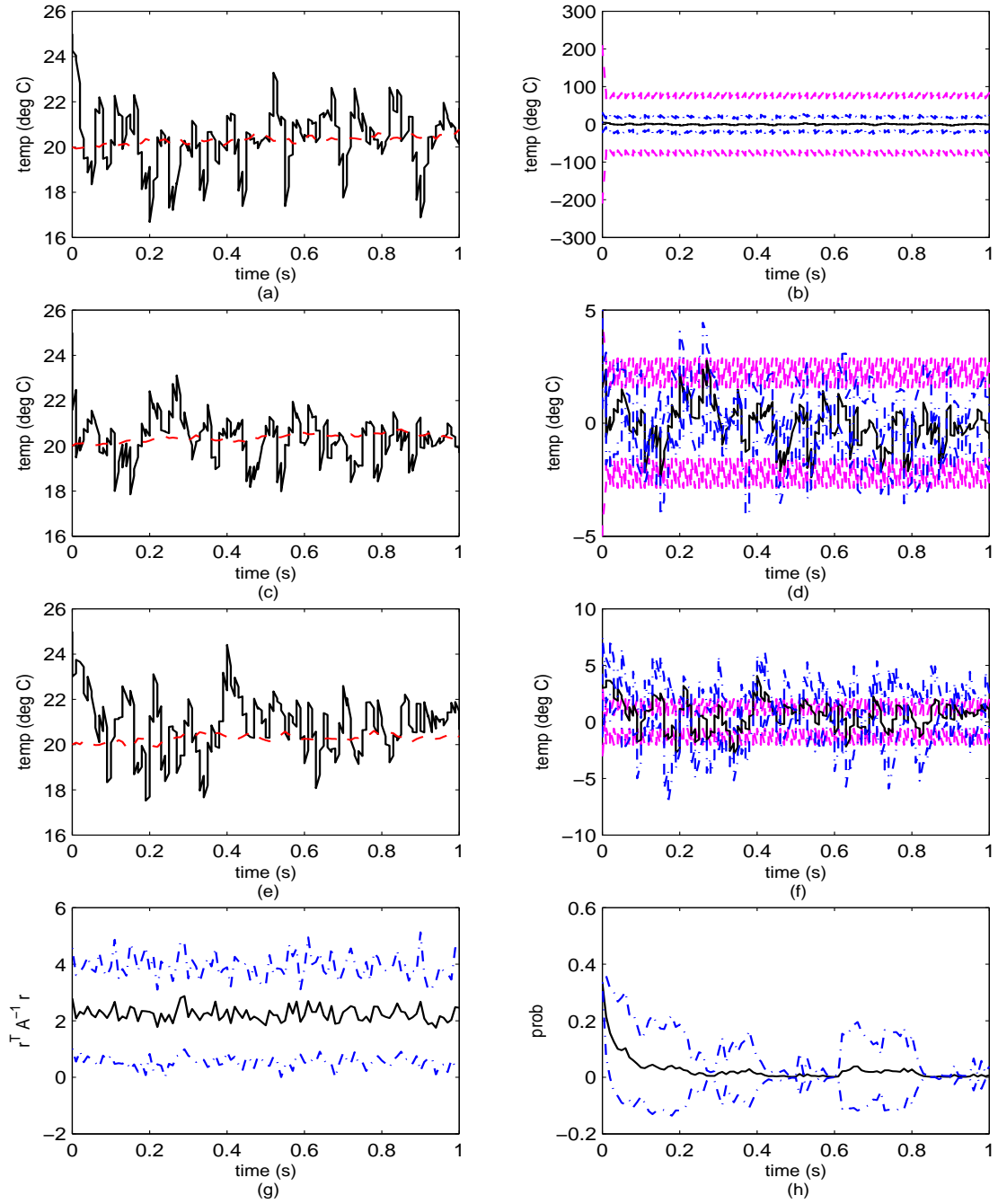


Figure 5.28 Simulation 2, Case 4 (P_0^{good}): Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

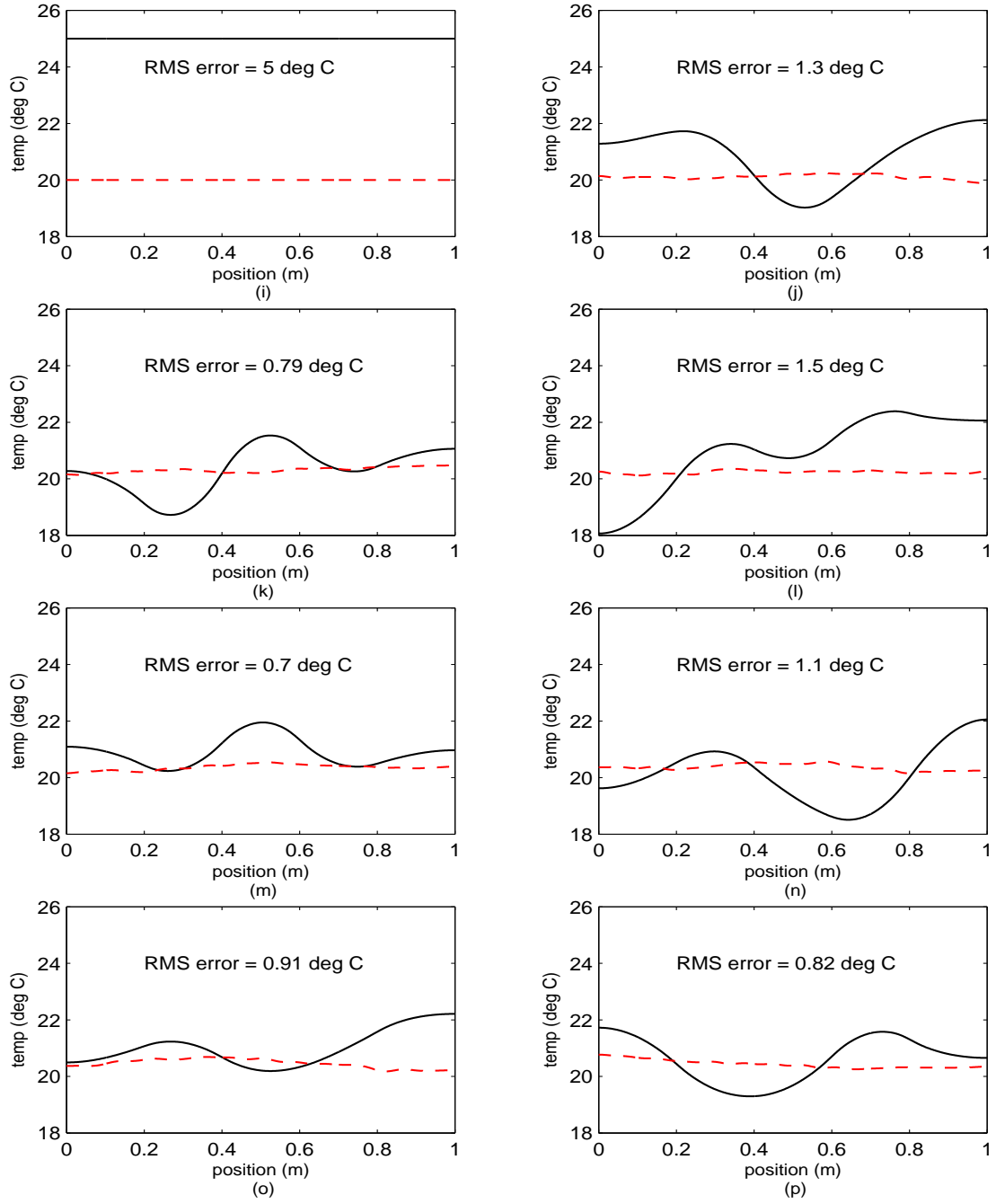


Figure 5.28 Simulation 2, Case 4 (P_0^{good}): Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

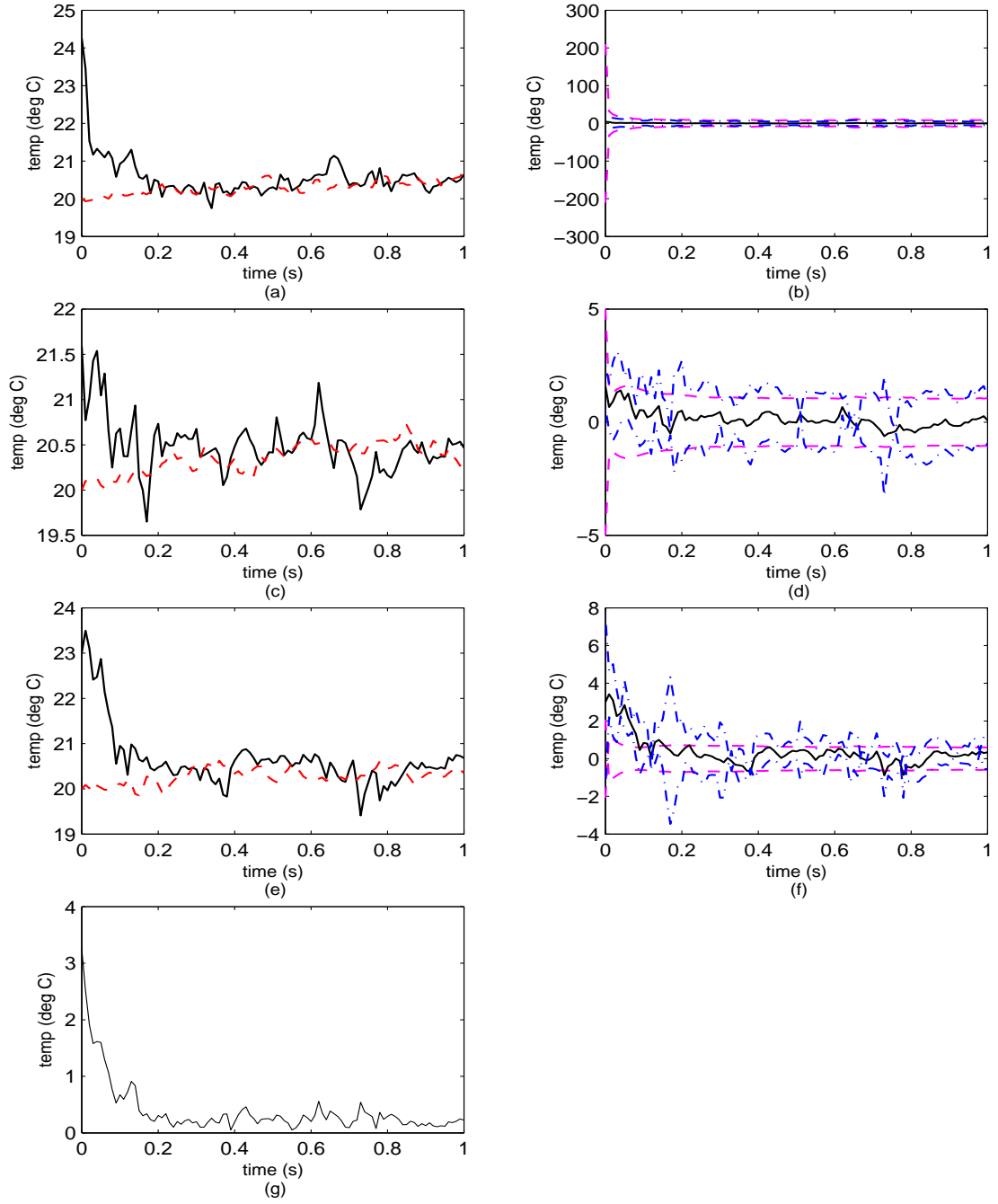


Figure 5.29 Simulation 2, Case 4 (P_0^{good}): Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

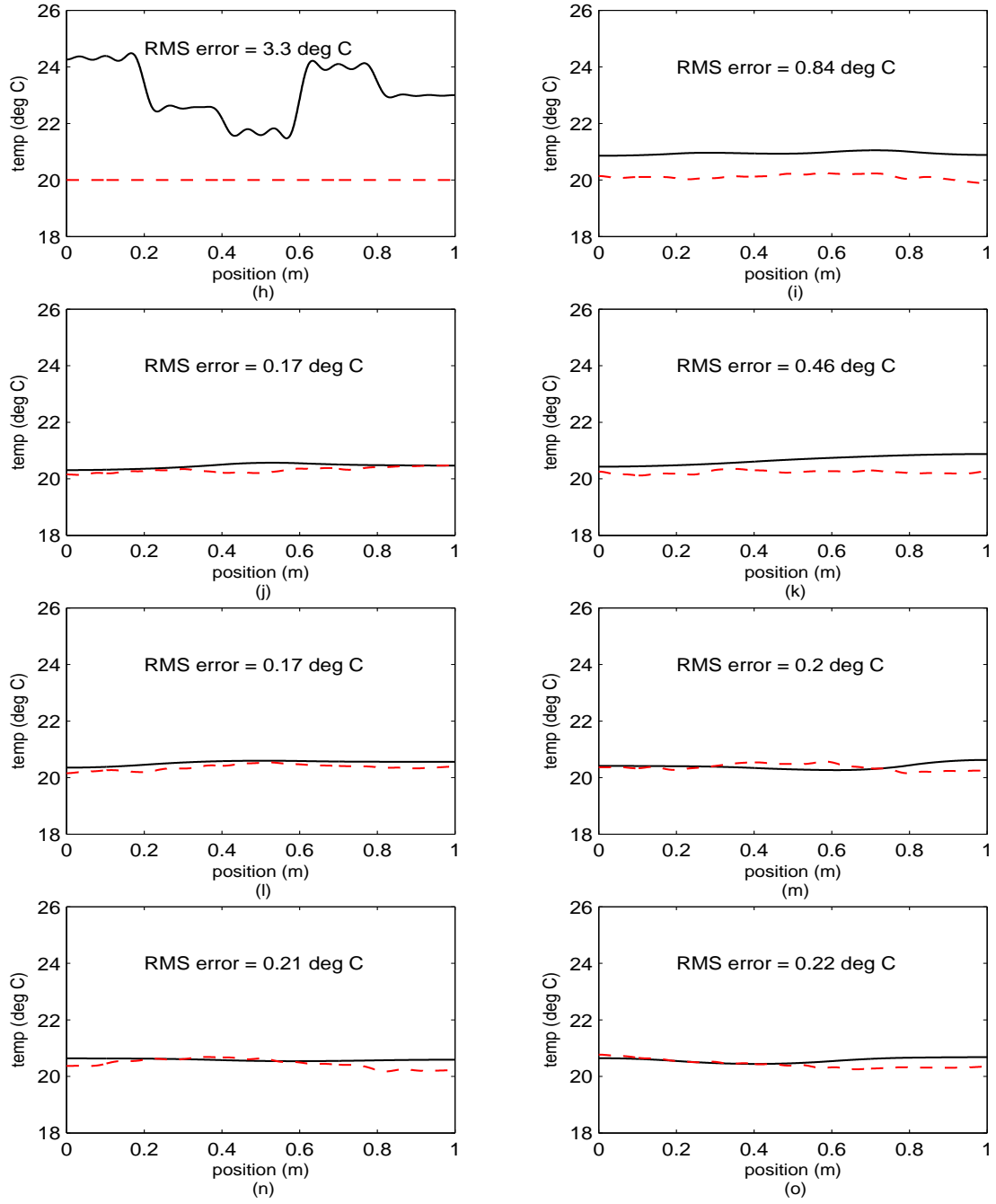


Figure 5.29 Simulation 2, Case 4 (P_0^{good}): Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

Blended estimates for all four cases. Since we have designed these experiments with full knowledge of what the truth is, the blended filter results are often inferior to the filter designed to match the simulated truth model fully. That is true here, as can be seen by comparing the figures for the second elemental filters to those of the blended filters. For example, compare Figures 5.19 and 5.21 or Figures 5.27 and 5.29. The blended result is biased because every elemental filter contributes to the blended solution. This bias is attributed to the fact that we have set a minimum threshold for the hypothesis conditional probabilities⁸ and thus, even filters based on completely mismatched models receive a nonzero weighting. We could of course choose to blend only those elemental filter estimates which exceed this threshold; that adjustment was not pursued in this research.

Finally, we draw your attention to another consequence of improperly setting the initial state covariance. Compare blended plots (h) for the Case 1 in Figure 5.17 and Case 2 in Figure 5.21, or for an even better visual, compare blended plots (h) for the Case 3 in Figure 5.25 and Case 4 in Figure 5.29. Note that for $P_0^{\text{good}} = 25$, the RMS error has been reduced from 5 to only 2.5 after just one measurement update; whereas for $P_0^{\text{poor}} = 1$, the RMS error has only been marginally reduced by about 5%. This is completely attributable to the choices for the initial state covariance. (As a side note, we can also determine the number of sensor segments used to gather the temperature data from the rod by noting the four transitions in Figures 5.21(h) and 5.29(h).)

5.4 Simulation 3

While the first simulation demonstrated how we could obtain an order-of-magnitude estimate of Q using an MMAE filter bank populated with five elemental filters, this simulation shows how we can get an accurate estimate of the measurement-corruption noise covariance, \mathbf{R} , using another bank of five elemental

⁸See Section 2.4.6 for a discussion on lower bounding the hypothesis conditional probabilities.

Case	Filter	Q	R	\hat{x}_0	P_0
1	1	Q_{true}	$\frac{1}{25}R_{\text{true}} = 0.2$	25	25
	2		$\frac{1}{5}R_{\text{true}} = 1$		
	3		$R_{\text{true}} = 5$		
	4		$5R_{\text{true}} = 25$		
	5		$25R_{\text{true}} = 125$		
2	1	Q_{true}	$\frac{2}{25}R_{\text{true}} = 0.4$	25	25
	2		$\frac{2}{5}R_{\text{true}} = 2$		
	3		$2R_{\text{true}} = 10$		
	4		$10R_{\text{true}} = 50$		
	5		$50R_{\text{true}} = 250$		

Table 5.7 Simulation 3: Elemental Filter Parameters

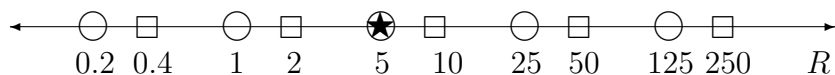


Figure 5.30 Simulation 3: Measurement-corruption noise covariance. Legend: \bigcirc elemental filters for Case 1; \square elemental filters for Case 2; \star true parameter within the filter bank. (The filter spacing is nonlinear, but appears linear for illustration purposes.)

filters. The first simulation demonstrated that the MMAE had trouble distinguishing among “closely” spaced filters in terms of the Q parameter — the best results were obtained for a discretization level of 100. Discretization of the R parameter does not suffer from this malady and hence the focus shifts to having enough filters to cover the range of possibilities. Table 5.7 presents the pertinent elemental filter design parameters for two cases of interest. In the first case, we place the “center” filter at what we know is the value for R_{true} and in the second case, we allow for a slight upward offset from this position. Figure 5.30 gives a graphical display of the separation between the elemental filters for both cases together on a single axis.

Note that the assumed R values for the five filters are separated from their nearest neighbors by just a factor of five (versus one to two orders of magnitude for Q in the previous simulation). Even though we saw in Simulation 2 that the best

filter tends to overestimate the true dynamics noise strength for “fine” discretization levels or for low Q/R ratio, in this simulation, discrimination of the five elemental filters did not depend strongly on the Q used. We obtained essentially the same results for $2Q$ and $5Q$, while at $10Q$ the state estimation performance started to degrade. However, measurement noise covariance estimation was still good. Only the results for truth Q are shown in the following figures for the filter bank centered on true R and the filter bank offset from R_{true} .

As we saw in the previous simulation, a poor choice for P_0 can have disastrous effects on the initial response of the system to new measurements. When P_0 is too small, the response of the filter to new measurements is slowed. This effect is much the same as overestimating the true R . Thus, only the results for an adequate initial state covariance is given.

As expected, the hypothesis conditional probability for elemental filter 3 tends towards one in both filter banks, as shown in Figure 5.31 while the probability for the other elemental filters tends towards zero, but at a faster rate than seen in Simulation 1. The centered filter bank is perhaps unrealistic for a real world situation, but as we can see here, the results for Case 2 in which the true measurement-corruption noise covariance occurs between the assumed values of two filters in the bank is still quite good as shown in Figure 5.31(b). These trends were very similar for the other three values of Q tested; thus we primarily discuss the case in which we know what the true value of Q is. Additionally, simulation 1 showed us that we need only a good guess to achieve good results since varying Q slightly had no appreciable effects on performance.

The results for Case 1 shown in plot (a) of Figure 5.31 are to be expected since the third filter was designed using the foreknowledge of the truth model. Similarly, we anticipate that the two elemental filters closest to the true value of the noise covariance (an underestimate and an overestimate) would receive “all” of the probability flow for Case 2, as shown in plot (b). As we saw in simulation 1, a

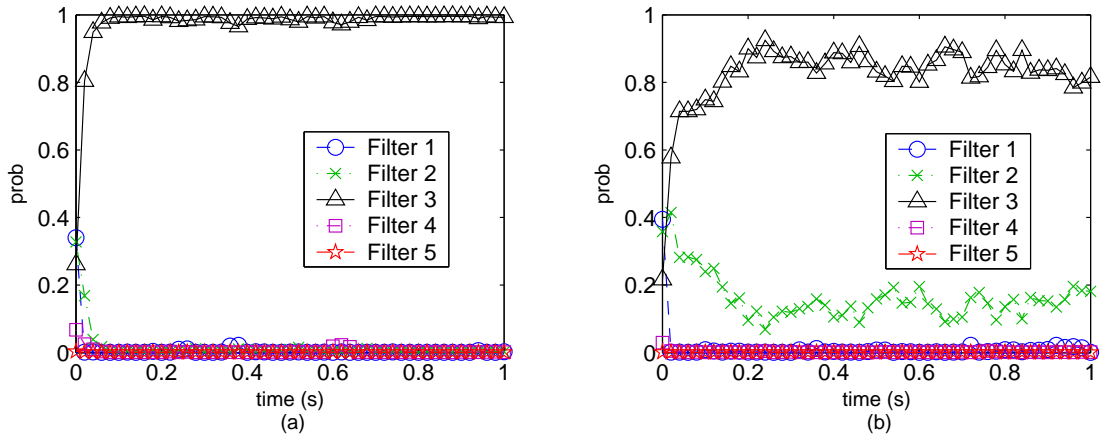


Figure 5.31 Simulation 3: Hypothesis conditional probability flow. (a) Case 1 filter bank: $\{\frac{1}{25}R_{\text{true}}, \frac{1}{5}R_{\text{true}}, R_{\text{true}}, 5R_{\text{true}}, 25R_{\text{true}}\}$. (b) Case 2 filter bank: $\{\frac{2}{25}R_{\text{true}}, \frac{2}{5}R_{\text{true}}, 2R_{\text{true}}, 10R_{\text{true}}, 50R_{\text{true}}\}$. To make these plots more legible, only the mean hypothesis conditional probabilities for times $\{t_0, t_2, t_4 \dots, t_{100}\}$ are displayed.

filter that slightly overestimates the expected covariance is favored over one that underestimates the covariance. A filter based on an underestimate for the covariance has its assumptions violated on a regular basis since the real world noise covariance exceeds that which was programmed into the elemental filter. Thus, the third elemental filter receives the bulk of the probability while the second elemental filter often underestimates the true covariance, it is correct often enough so that it receives a small (but noteworthy) share of the probability. Note that if we try to use an even larger overestimate for the measurement-corruption noise covariance, then we end up with an elemental filter that “thinks” that the measurements are so sloppy as to be relatively worthless as compared to the quality of the dynamics model. This effect can be seen clearly in the probability calculation (given in Section 2.3.3.3) — that is, the probability that the hypothesis is correct is inversely proportional to the square-root of the determinant of the measurement-corruption noise covariance, i.e., an excessively large covariance can yield a very small probability due to this scaling term. For a reasonable overestimate, this scaling is balanced by the exponential portion of the probability density function.

Case 1: Adapting to an unknown R_{true} using a filter bank centered on R_{true} .

It comes as no surprise that, when we match an elemental-filter-assumed parameter value to the real-world parameter value, the MMAE performs very well. We begin our analysis by observing the probability flow in Figure 5.31(a). Then, we investigate the (h) plots for the individual elemental filters for this case in Figures 5.32 through 5.36 to see how much the hypothesis conditional probabilities varied over the course of the simulation. For this case, only elemental filter 3 seems to have been based on a good hypothesis of the true R ; this is not surprising. If we look at plot (g) in conjunction with plot (h), we can see that the likelihood quotient gives a reliable account of which filter matches the simulated world the best. When R_{filter} is smaller than R_{true} , the likelihood quotient is greater than the expected $M = 5$ that is exhibited by a filter for the best hypothesis. As the assumed value for R increases, the likelihood quotient decreases. As shown in the development presented in the introduction section for this chapter, the likelihood quotient is roughly equal to a ratio of the true R to the assumed R for the elemental filter. Hence, the results we have just seen in plots (g) are fairly predictable. As mentioned above, convergence is dictated by the Q/R ratio (which gives rise to the steady state gain K); hence the elemental filters with the smallest assumed R converged the quickest to the rod temperature.

Continuing our analysis for this case, we see that when the assumed R is significantly smaller than the true noise covariance, the dynamics model is essentially cast aside, i.e., the Kalman gain is very high. To support this assertion empirically, we inspect the behavior displayed in plots (a) through (f) in Figure 5.32 and to a lesser degree in Figure 5.33 since the assumed R is closer to the true R in the latter. Elemental filter 3 represents a balance between the dynamics and measurement models and is also the best elemental filter in the bank. The fourth and fifth elemental filters behave in the opposite manner, they “trust” the dynamics model more than they should and place less emphasis on the new measurements, which can be seen

by the slowly converging state estimates on Figure 5.35 and to a greater degree on Figure 5.36.

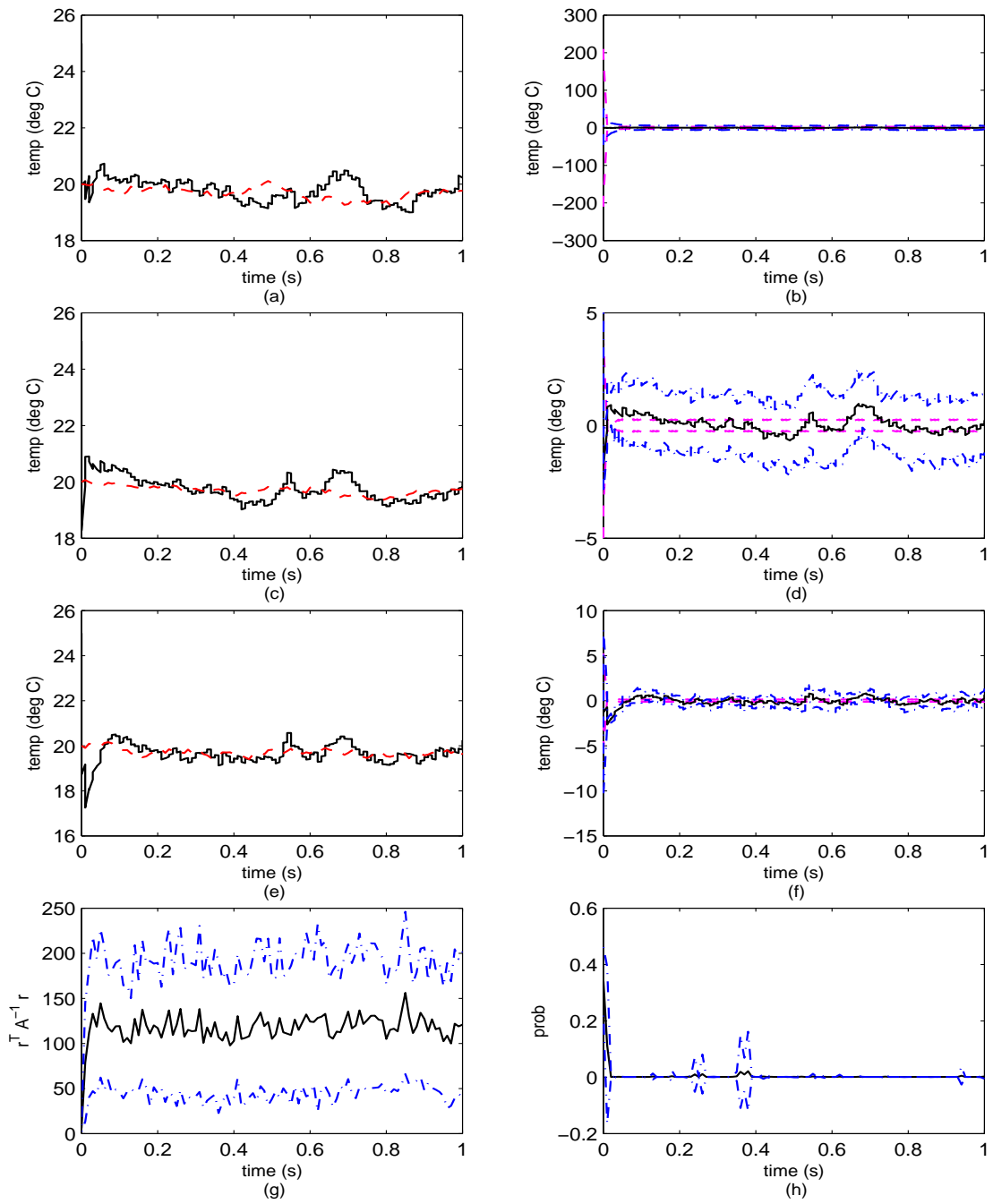


Figure 5.32 Simulation 3, Case 1: Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

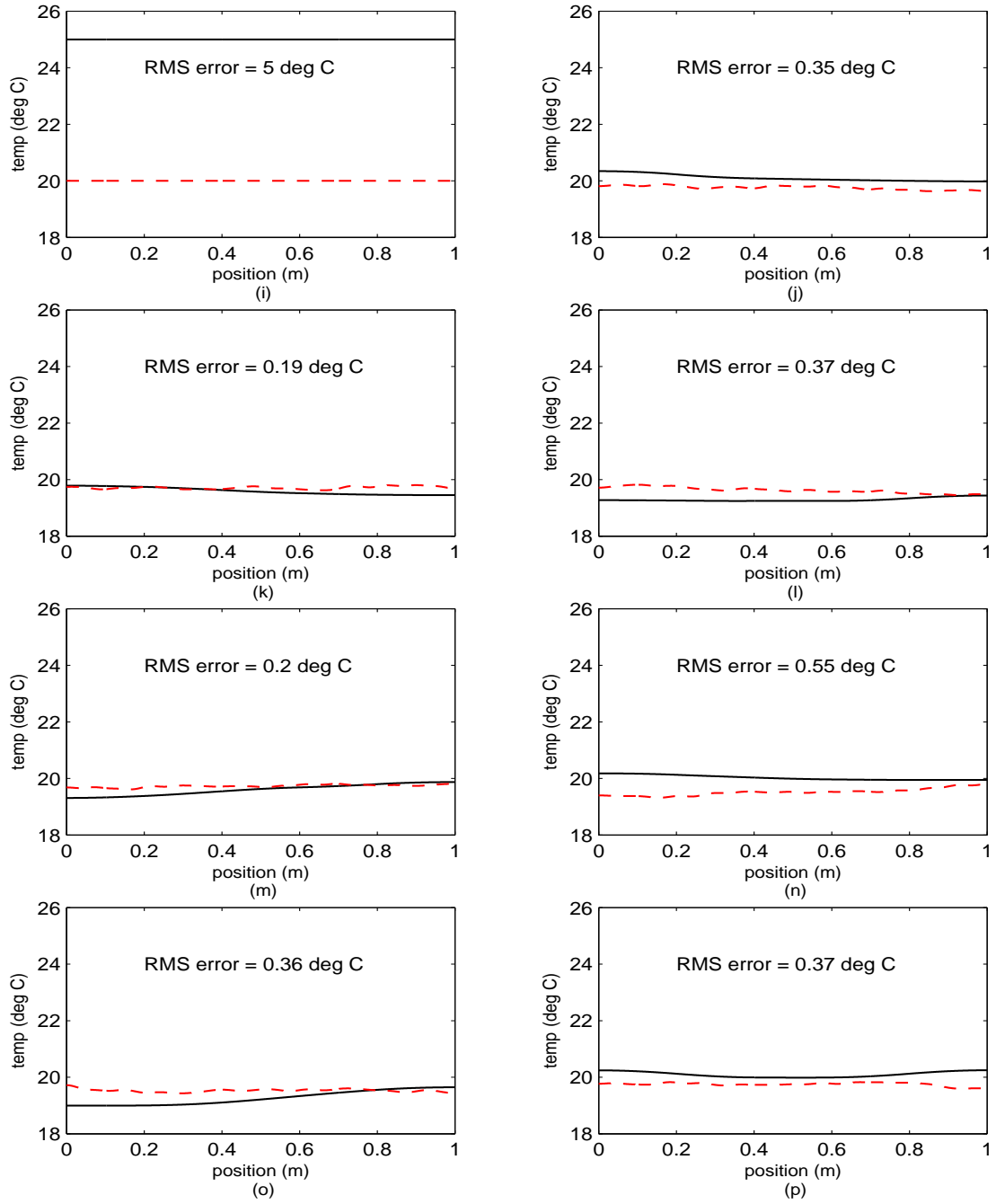


Figure 5.32 Simulation 3, Case 1: Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

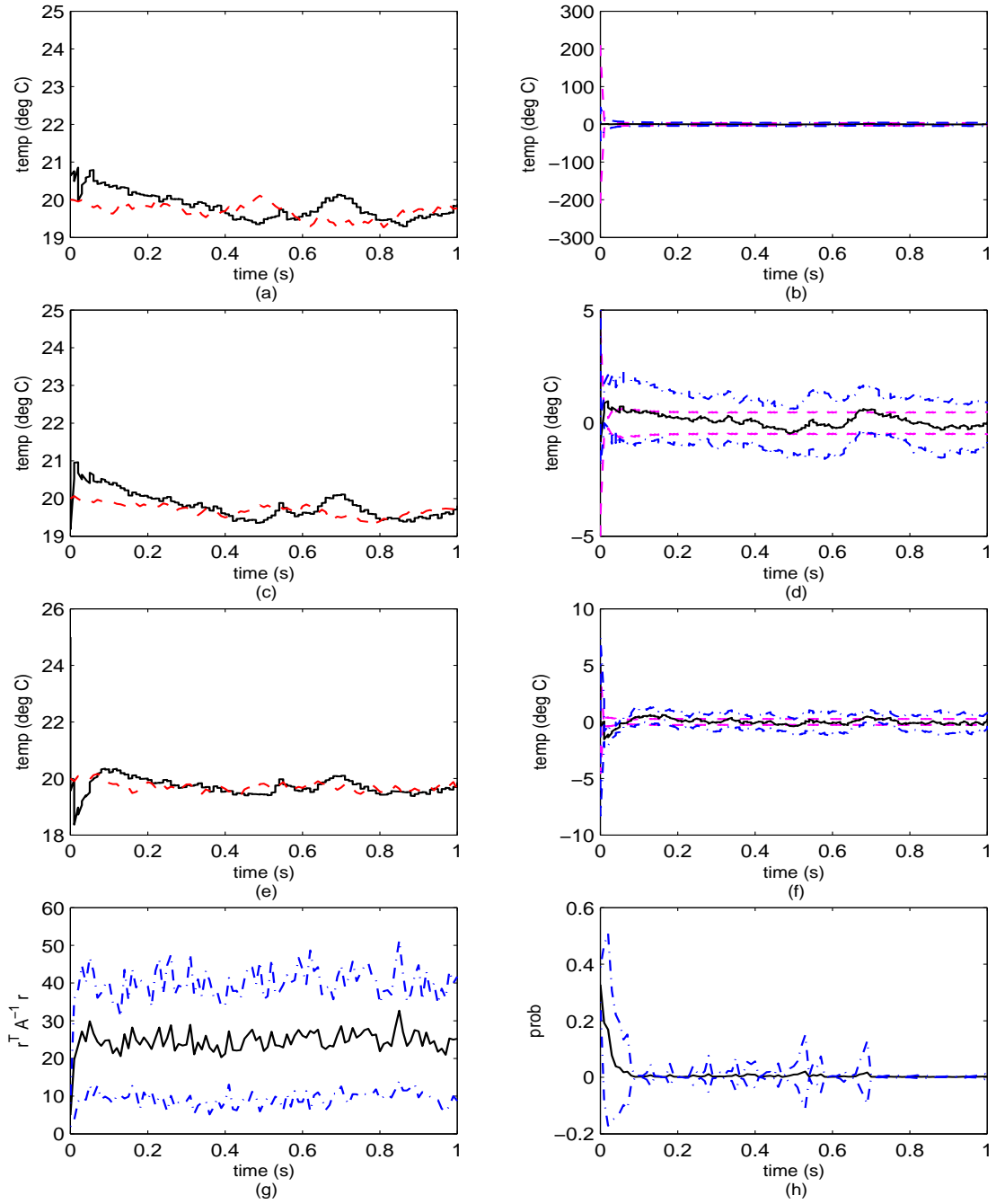


Figure 5.33 Simulation 3, Case 1: Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

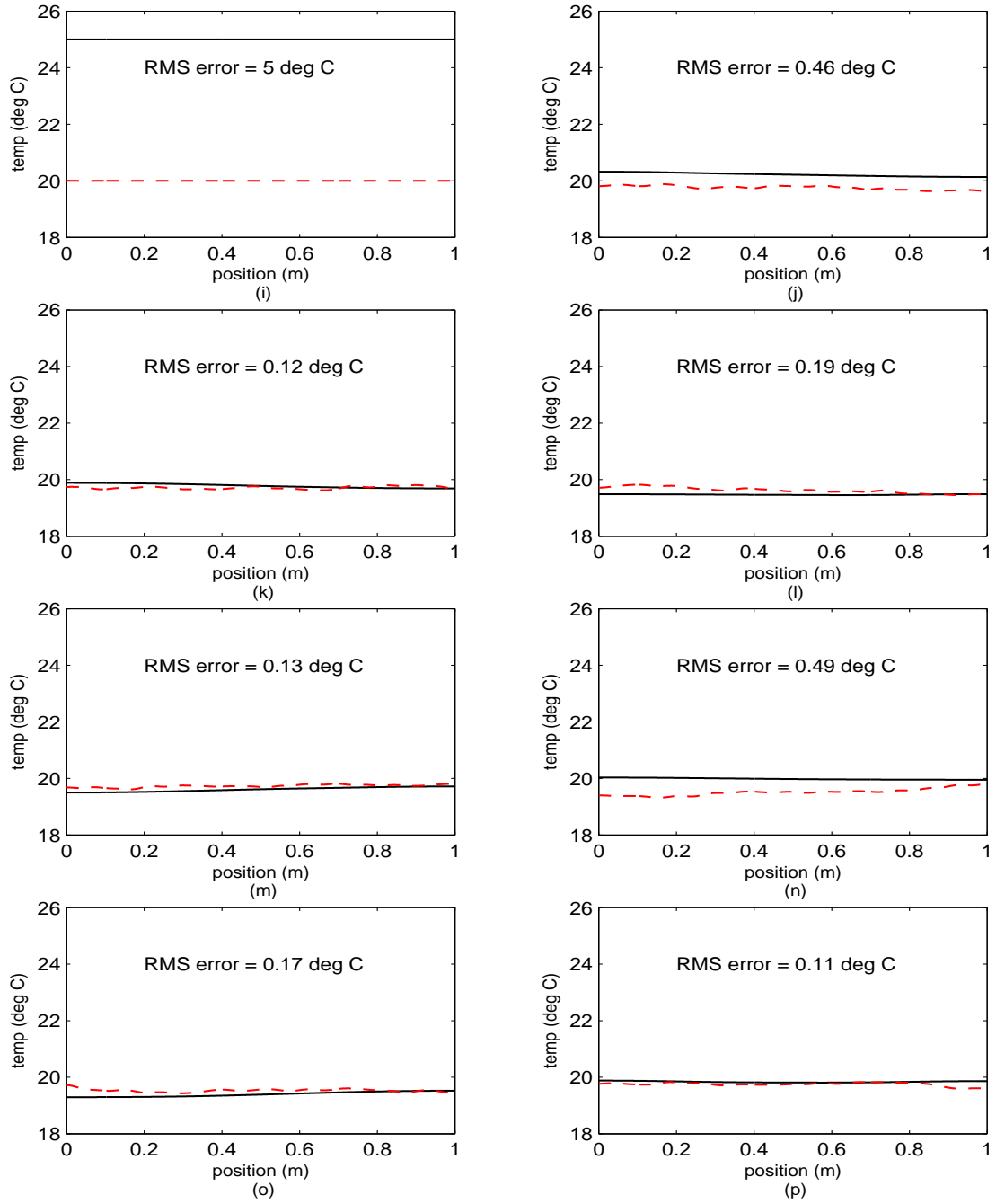


Figure 5.33 Simulation 3, Case 1: Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

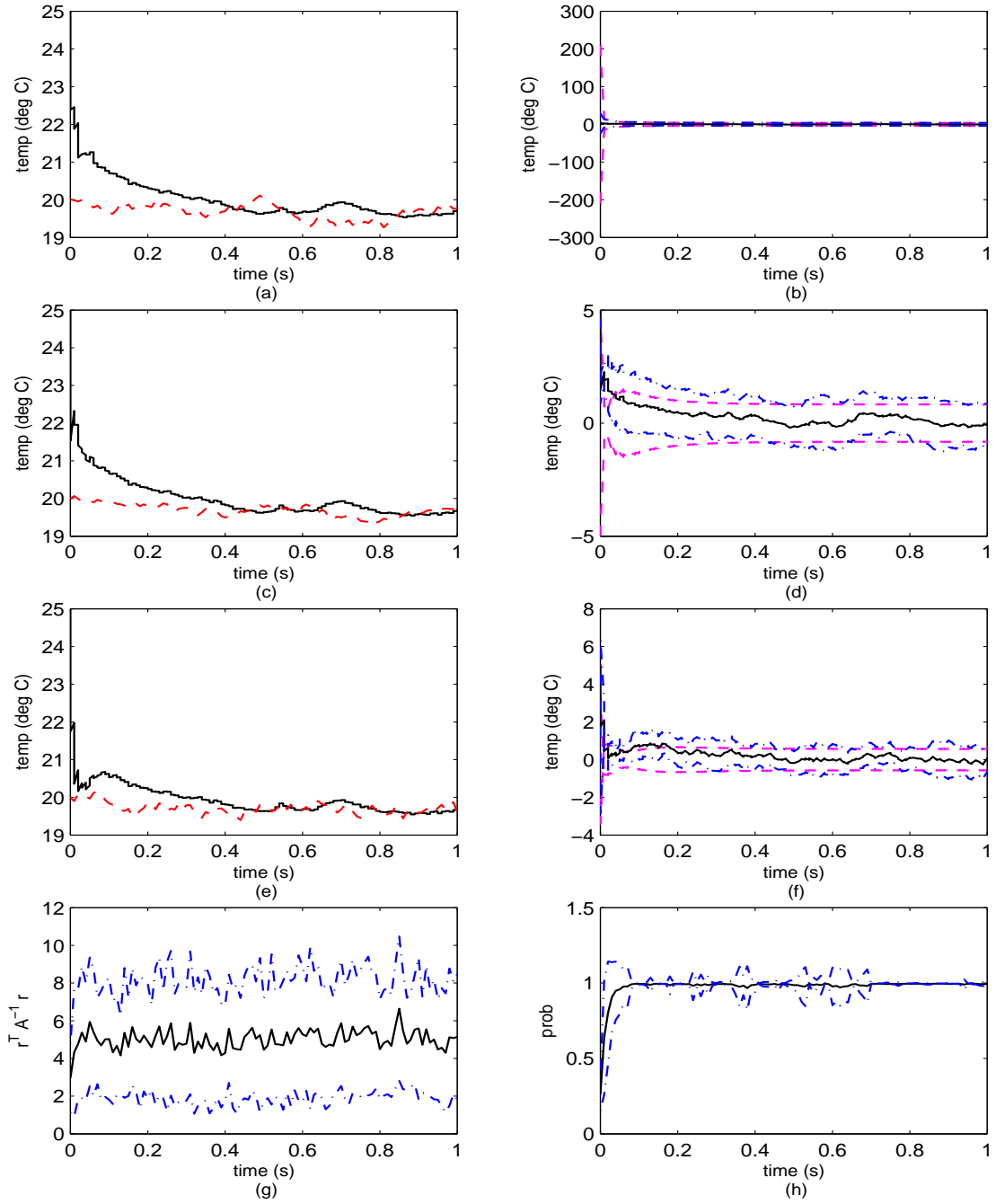


Figure 5.34 Simulation 3, Case 1: Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

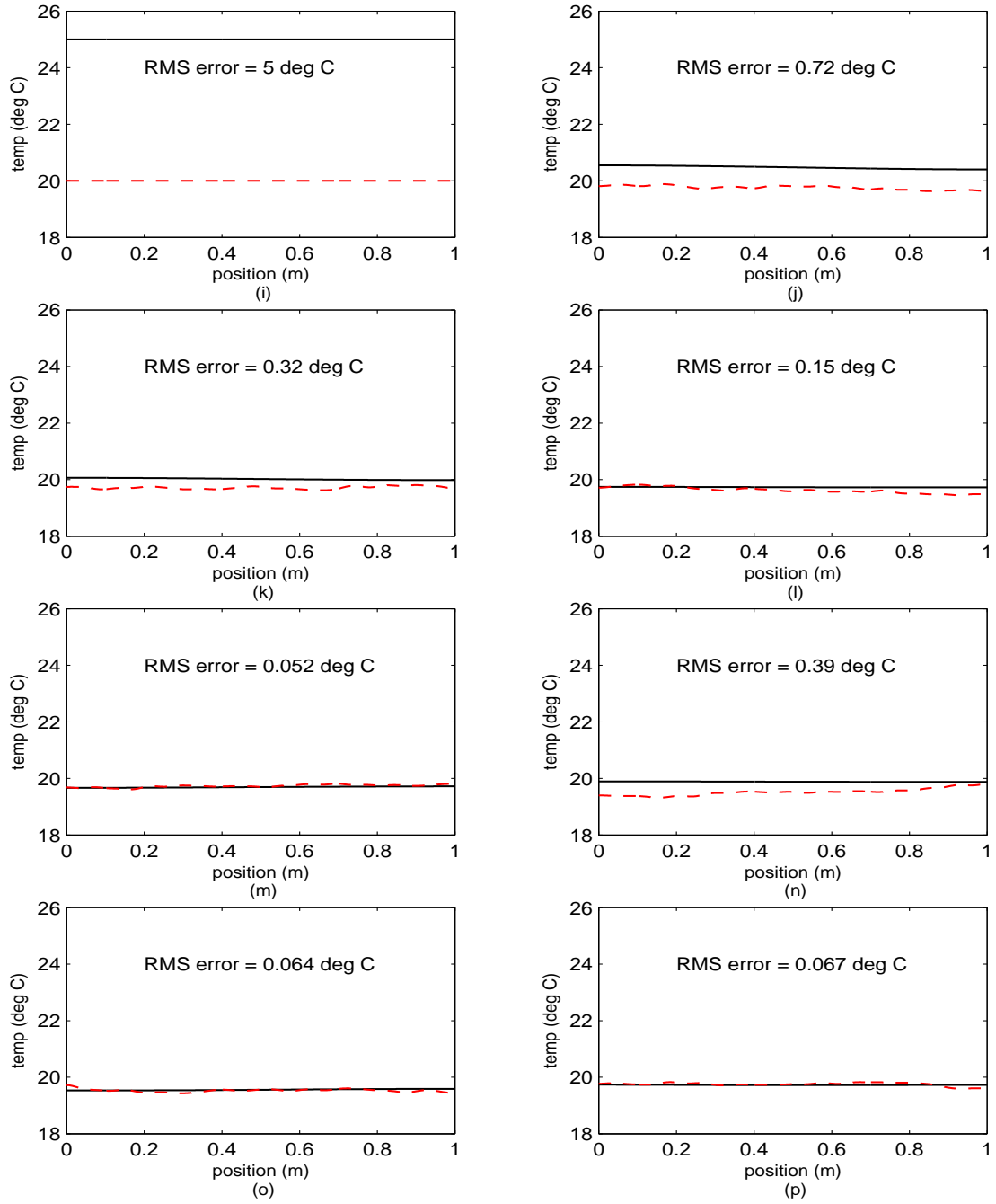


Figure 5.34 Simulation 3, Case 1: Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

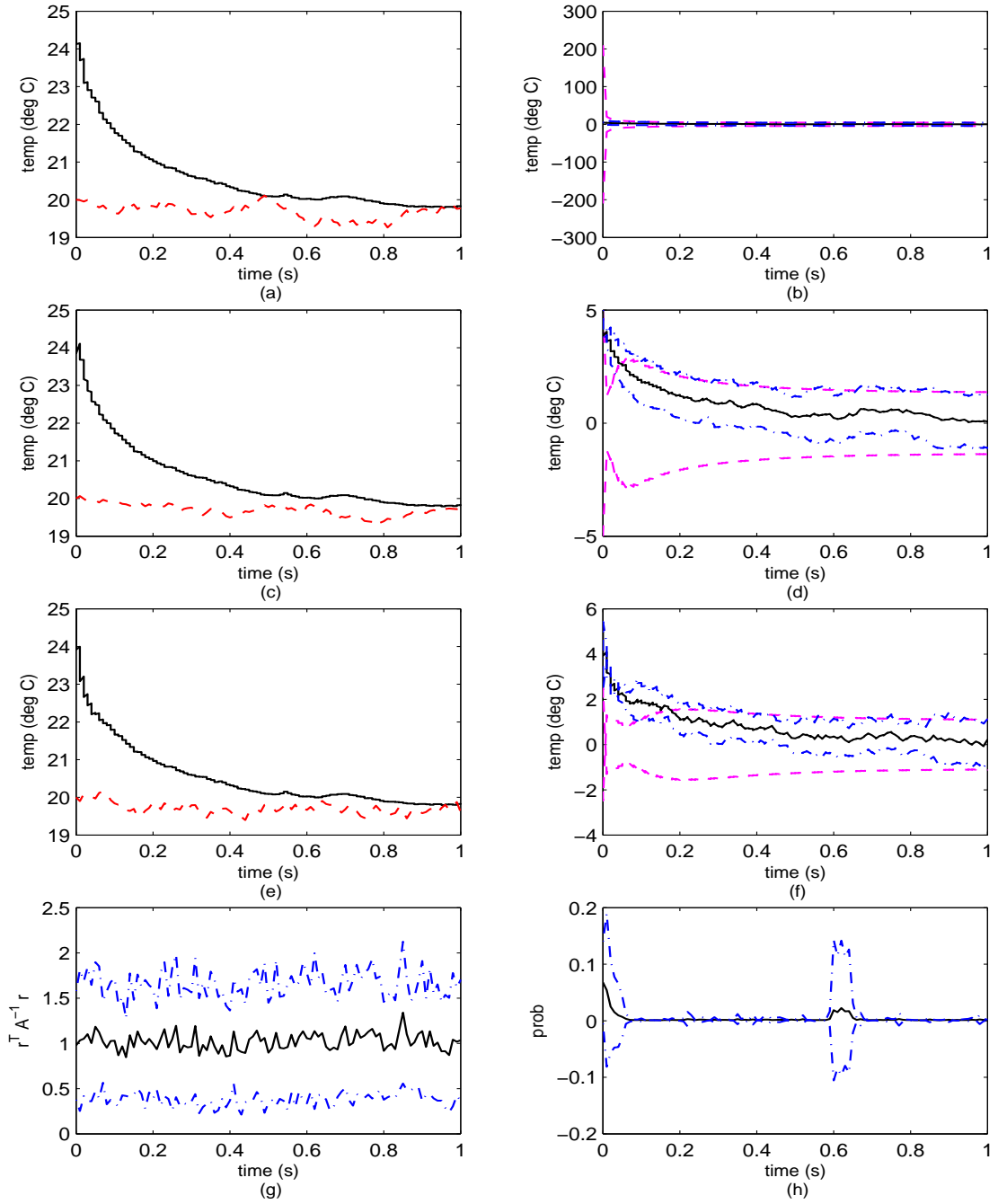


Figure 5.35 Simulation 3, Case 1: Elemental Filter 4. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

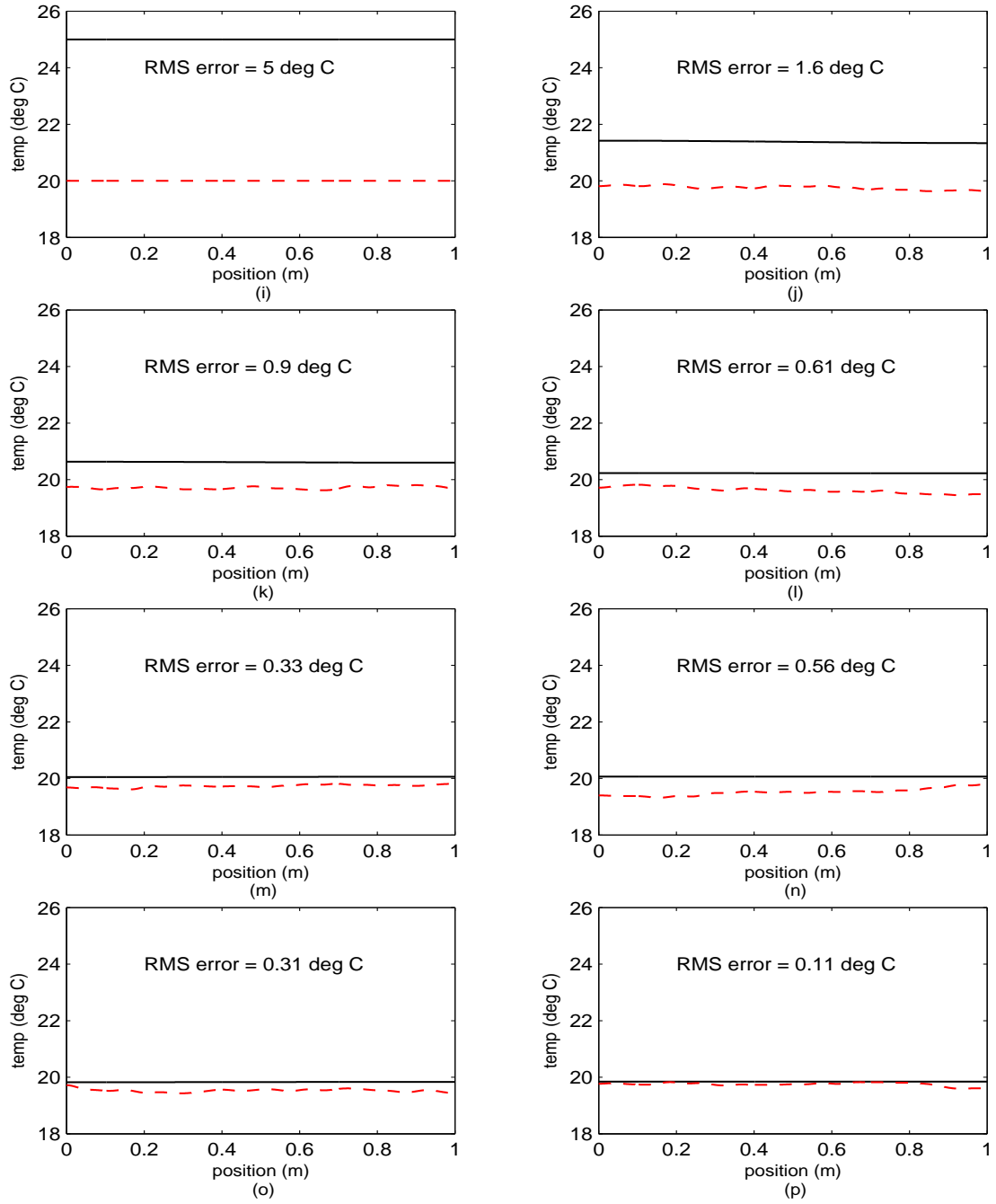


Figure 5.35 Simulation 3, Case 1: Elemental Filter 4 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

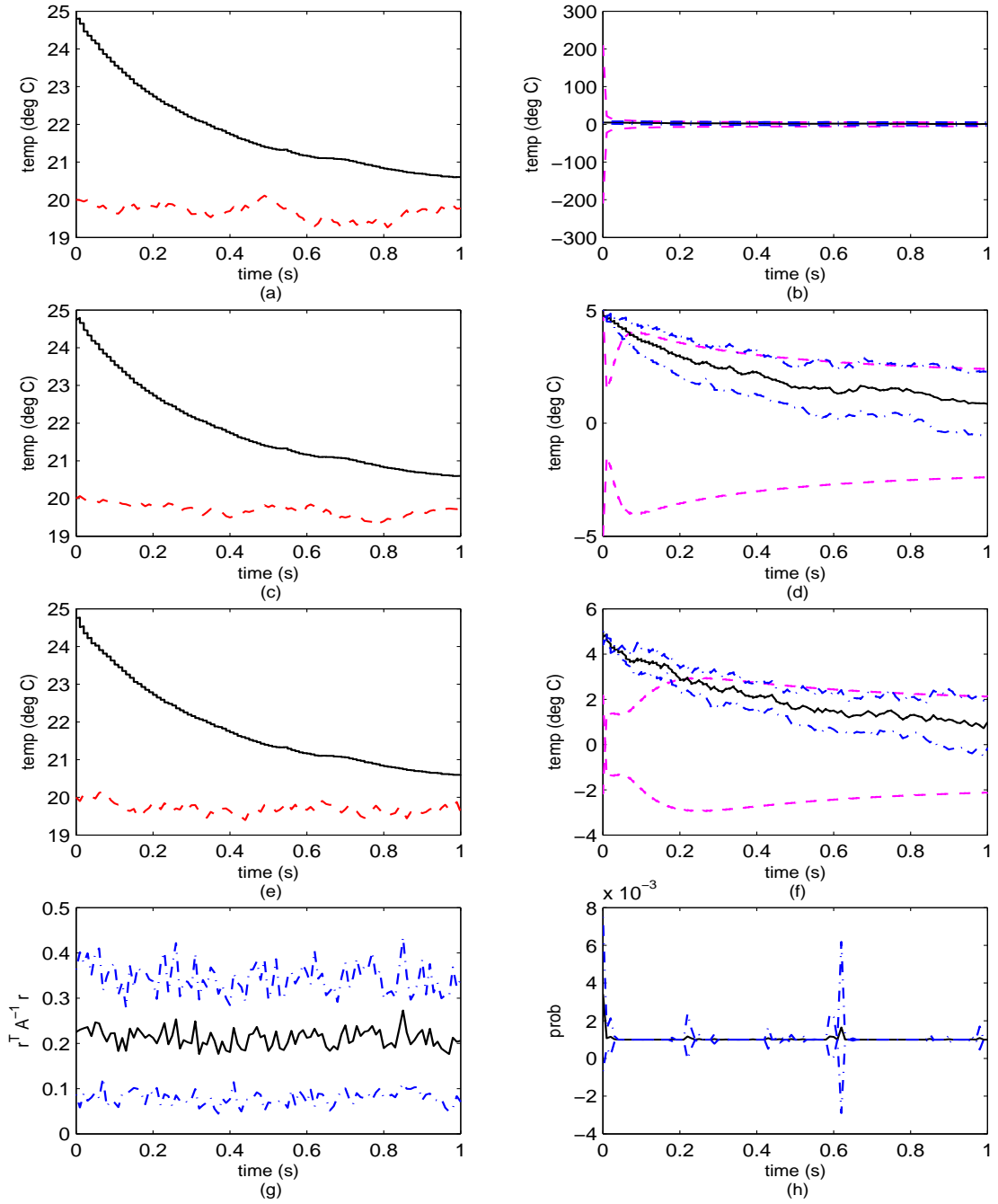


Figure 5.36 Simulation 3, Case 1: Elemental Filter 5. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

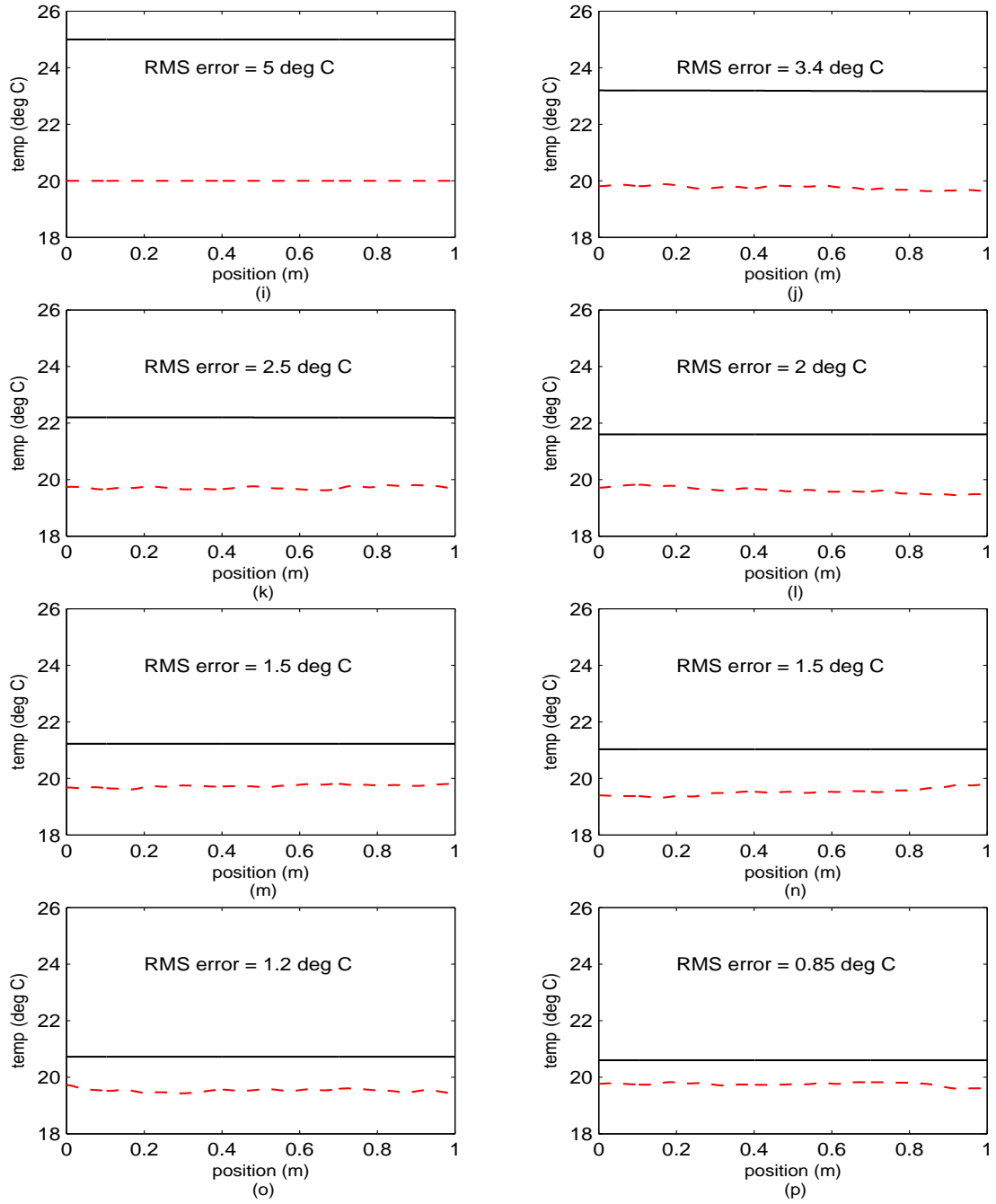


Figure 5.36 Simulation 3, Case 1: Elemental Filter 5 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

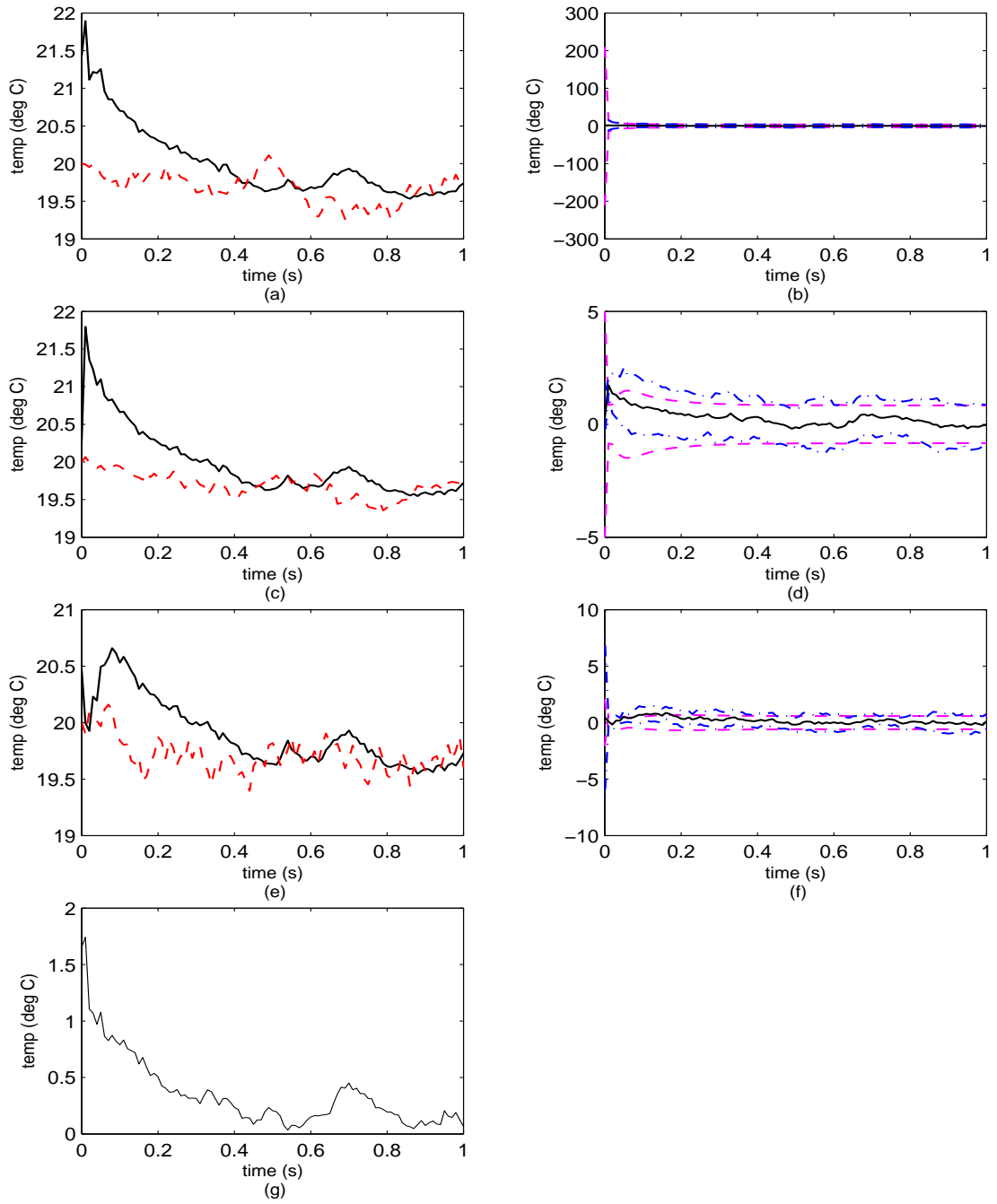


Figure 5.37 Simulation 3, Case 1: Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

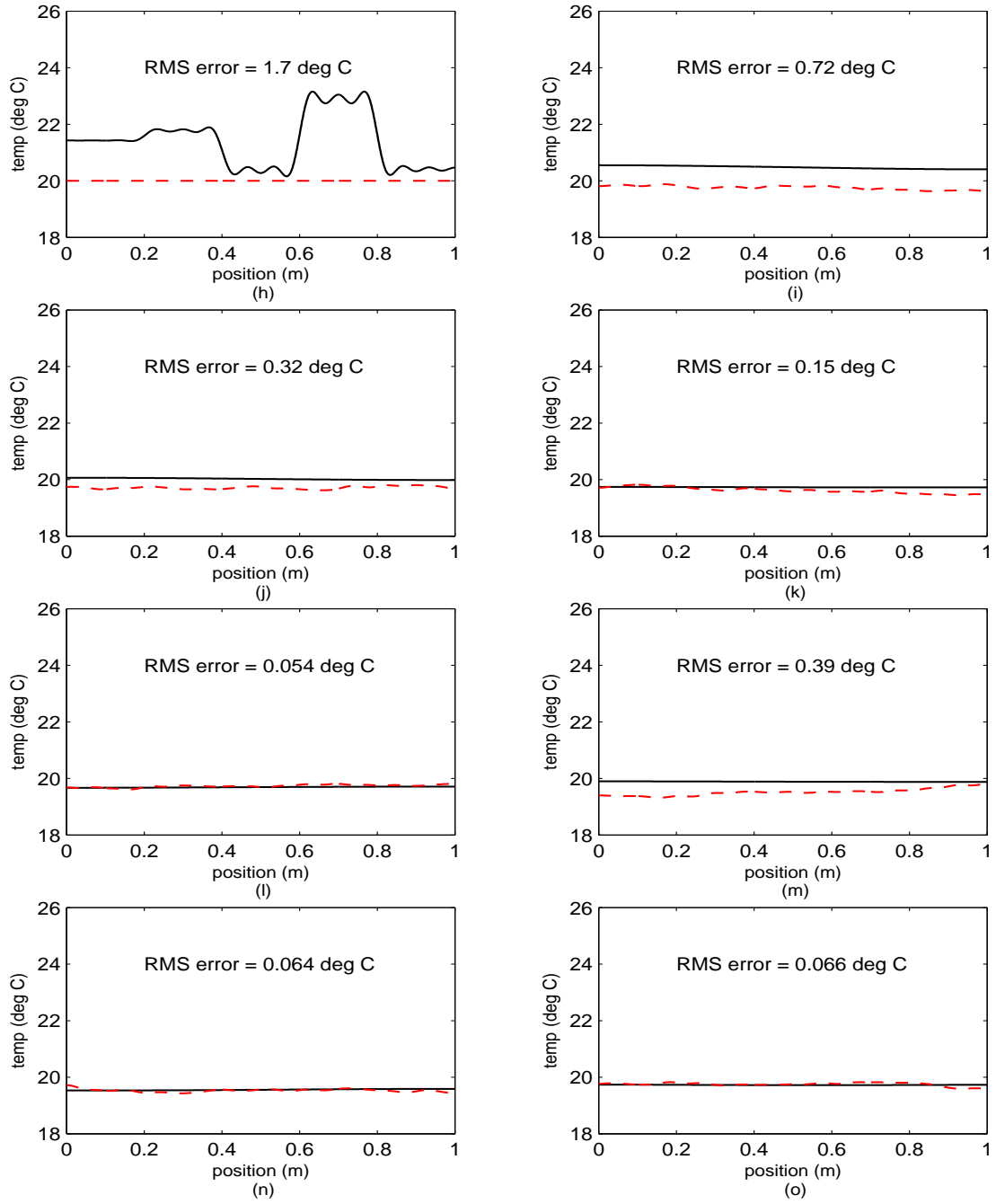


Figure 5.37 Simulation 3, Case 1: Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

Case 2: Adapting to an unknown R_{true} using a filter bank purposefully offset from R_{true} . As expected, the second and third elemental filters were deemed the most likely as seen in Figure 5.31(b). The hypotheses conditional probability histories in plot (h) of Figures 5.38 through 5.42, show that the probabilities for elemental filters 1, 4, and 5 are nearly zero in every Monte Carlo run, while the relative share of the probability varied considerably between elemental filters 2 and 3 as indicated by the rather large standard deviation evident in the mean plus and minus one sigma (dashed) lines on plots (h). Recall that, for each Monte Carlo run, a sample of the measurement-corruption noise process \mathbf{v} is drawn. (Over the course of the fifty runs, these samples are representative of a white Gaussian noise processes with covariance \mathbf{R} .) So, for a particular simulation run, if R is high — where R is the repeated eigenvalue of the matrix \mathbf{R} — then the MMAE would more heavily favor elemental filter 3, and if R is low, it more closely matched elemental filter 2. As anticipated, the mean probability flow to elemental filter 3 is greater than to elemental filter 2. An examination of plot (g) reveals that these two elemental filters are in consonance with the real-world conditions since their likelihood quotients are near $M = 5$; the likelihood quotient for elemental filter 3 ($R_3 = 2R_{\text{true}}$) is less than five (about two and a half), while the one for elemental filter 2 ($R_2 = 0.4R_{\text{true}}$) was too large (around 12). The rest of the elemental filters were poorly matched to the simulated real world and thus received essentially zero probability for the entire simulation, and their likelihood quotients in plot (g) were either too large (for filter 1), or too small (for filters 4 and 5). Note that the RMS error at time $t_i = 1$ sec is lowest for blended filter output [Figure 5.43)(o)], versus any of the elemental filters alone, even though none of the elemental filters were based on a correct value for \mathbf{R} . The blended filter output is due almost entirely to an effective blending of elemental filters 2 and 3. Hence, in a real-world environment, where we might have only incomplete and/or low quality information on the true noise environment, the blended estimates may prove to be the most useful.

A final comment. While it was entirely anticipated that elemental filter 3 would absorb all of the probability in Case 1, it is useful to note that even when the assumed value is slightly increased (in this case it was doubled), the elemental filter still matches quite well, as demonstrated in Case 2. Additionally, the elemental filter that “resides on the other side” of the true \mathbf{R} gathered the remainder of the probability, as previously noted.

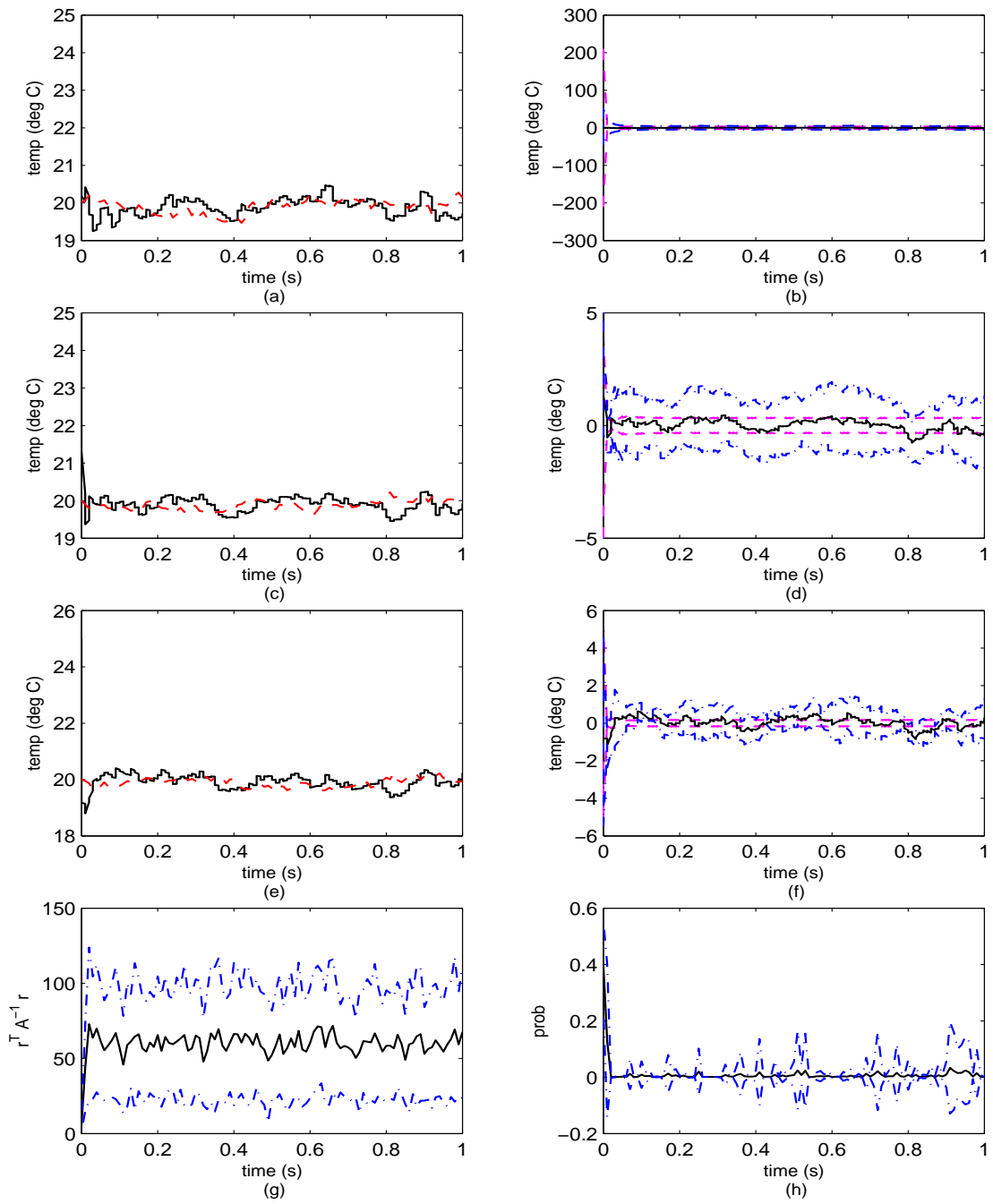


Figure 5.38 Simulation 3, Case 2: Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

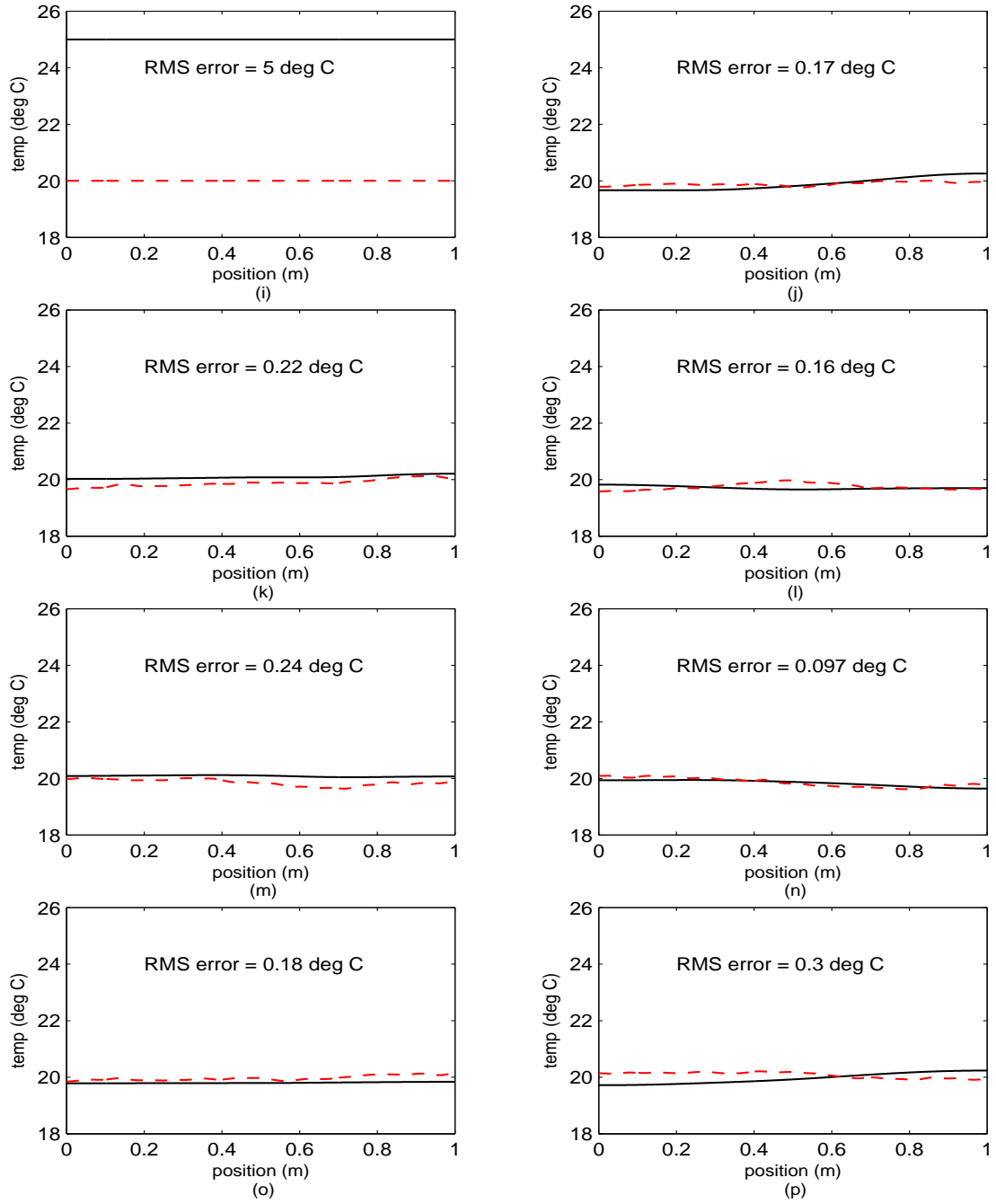


Figure 5.38 Simulation 3, Case 2: Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

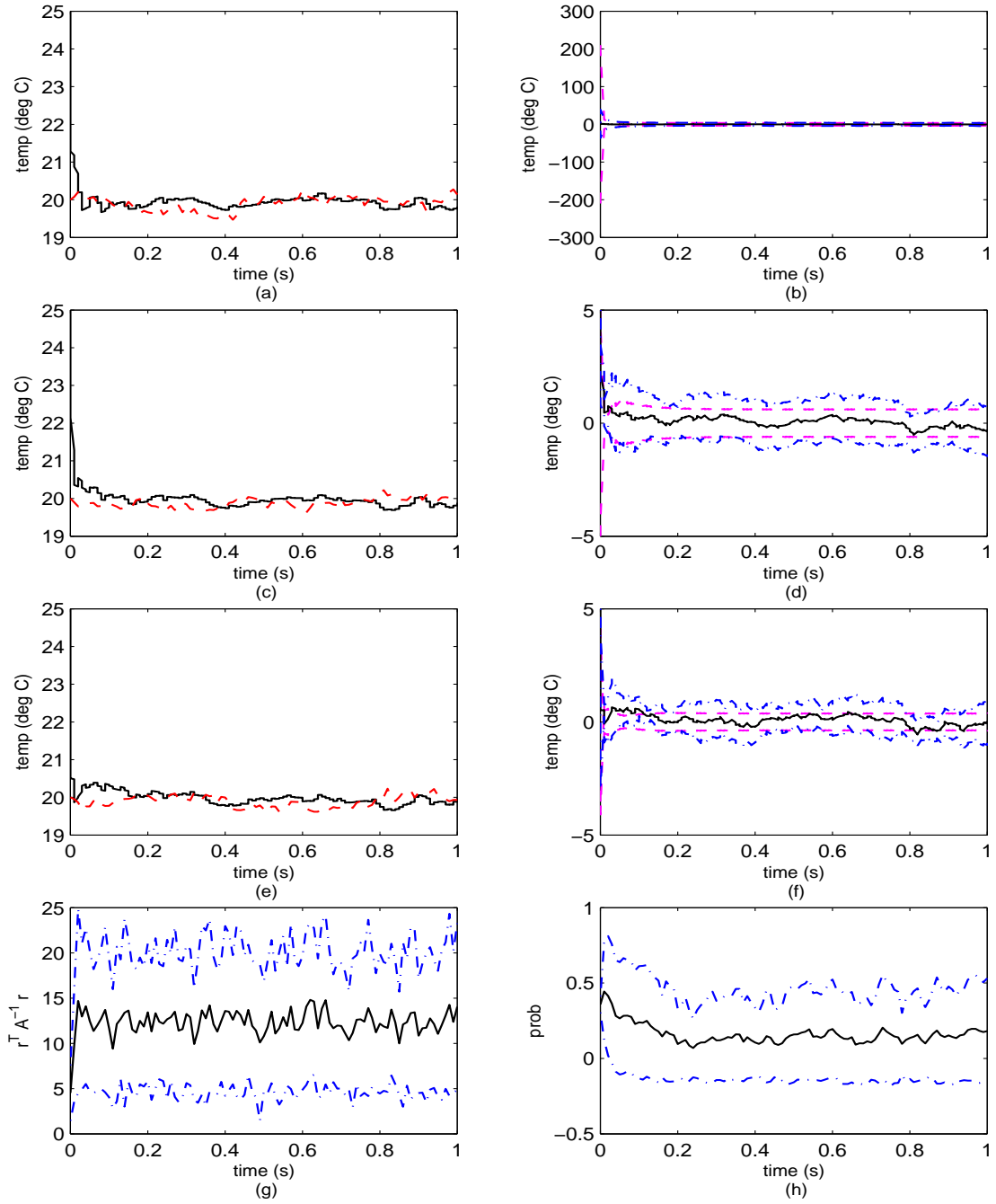


Figure 5.39 Simulation 3, Case 2: Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

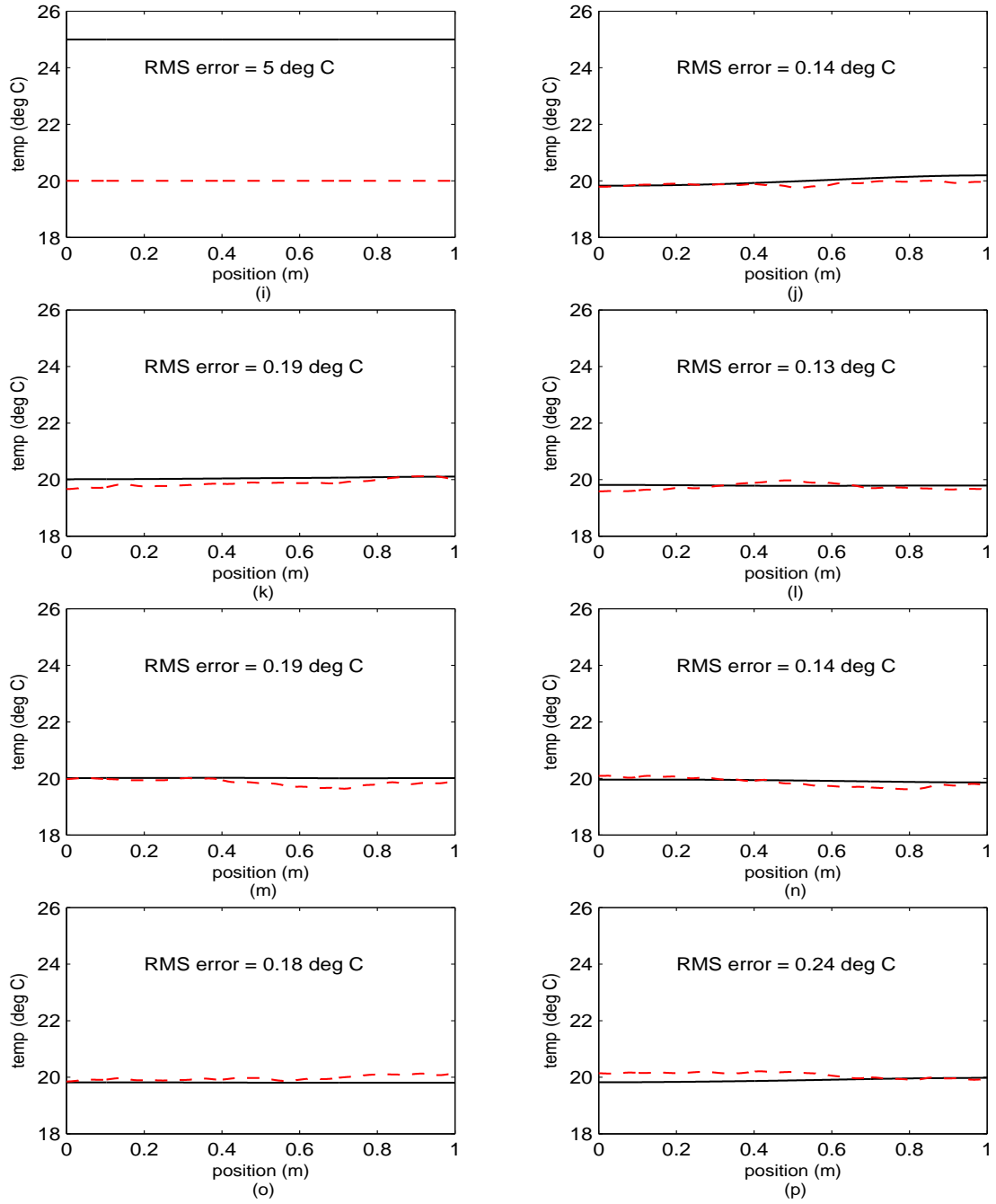


Figure 5.39 Simulation 3, Case 2: Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

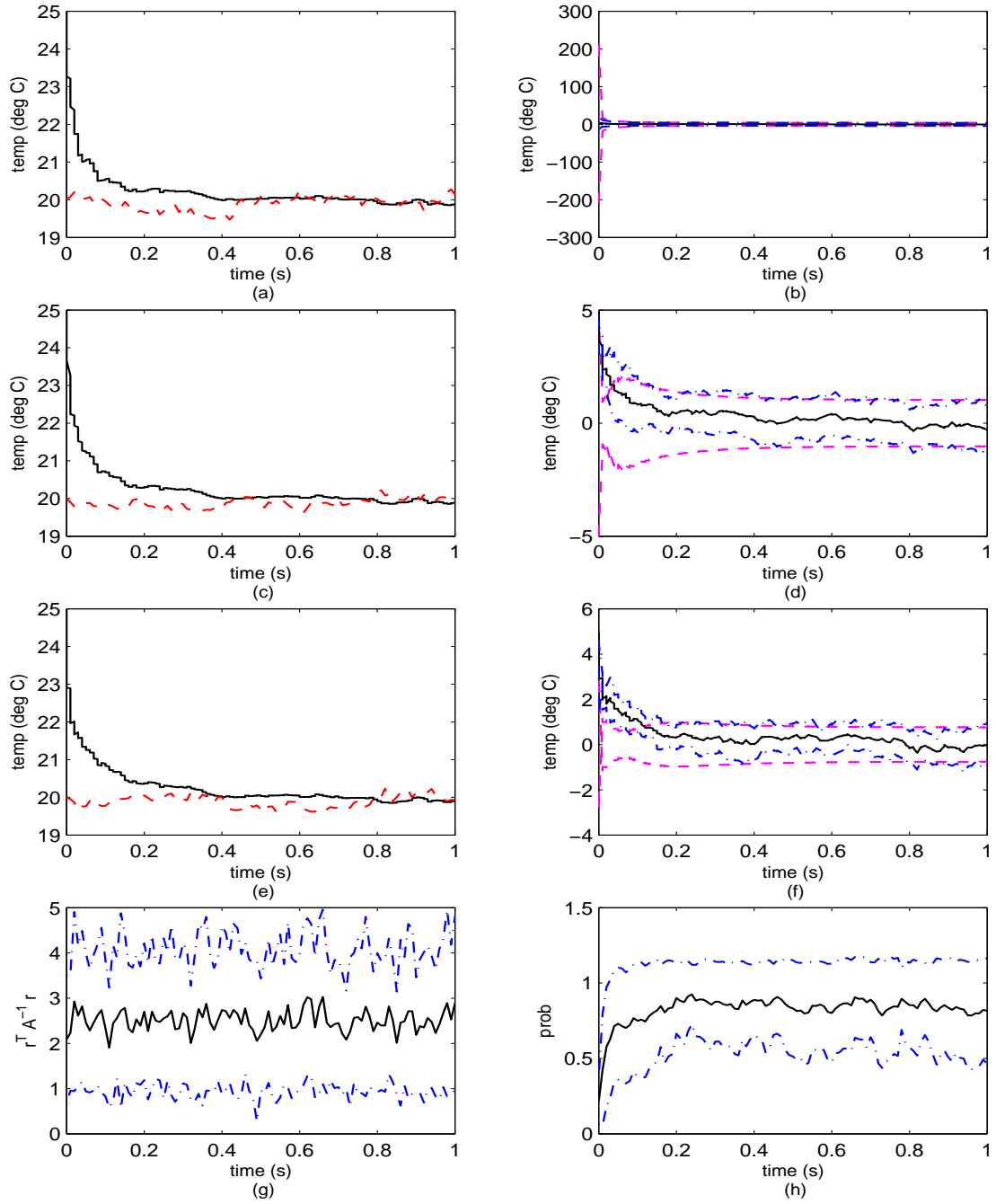


Figure 5.40 Simulation 3, Case 2: Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

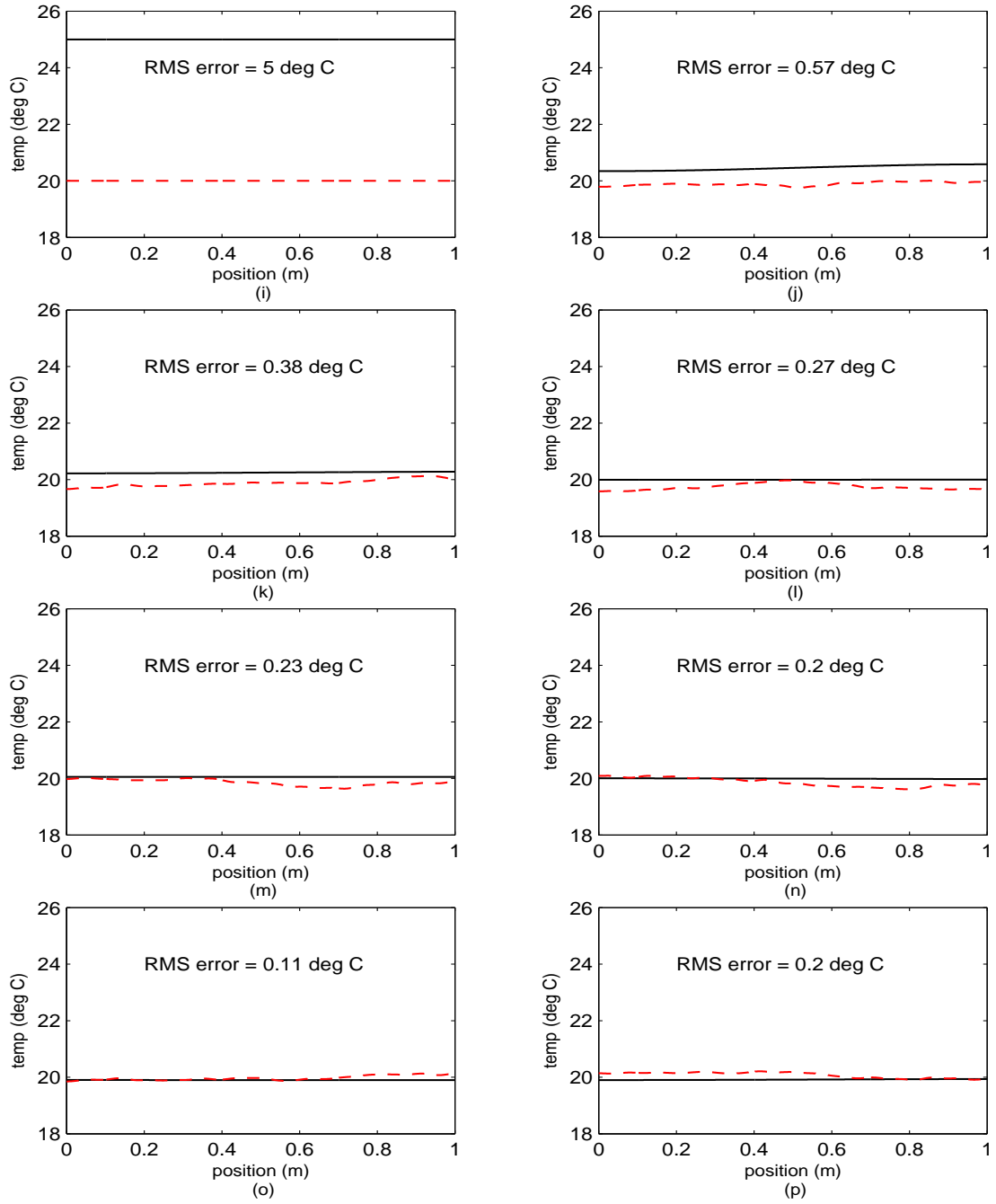


Figure 5.40 Simulation 3, Case 2: Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

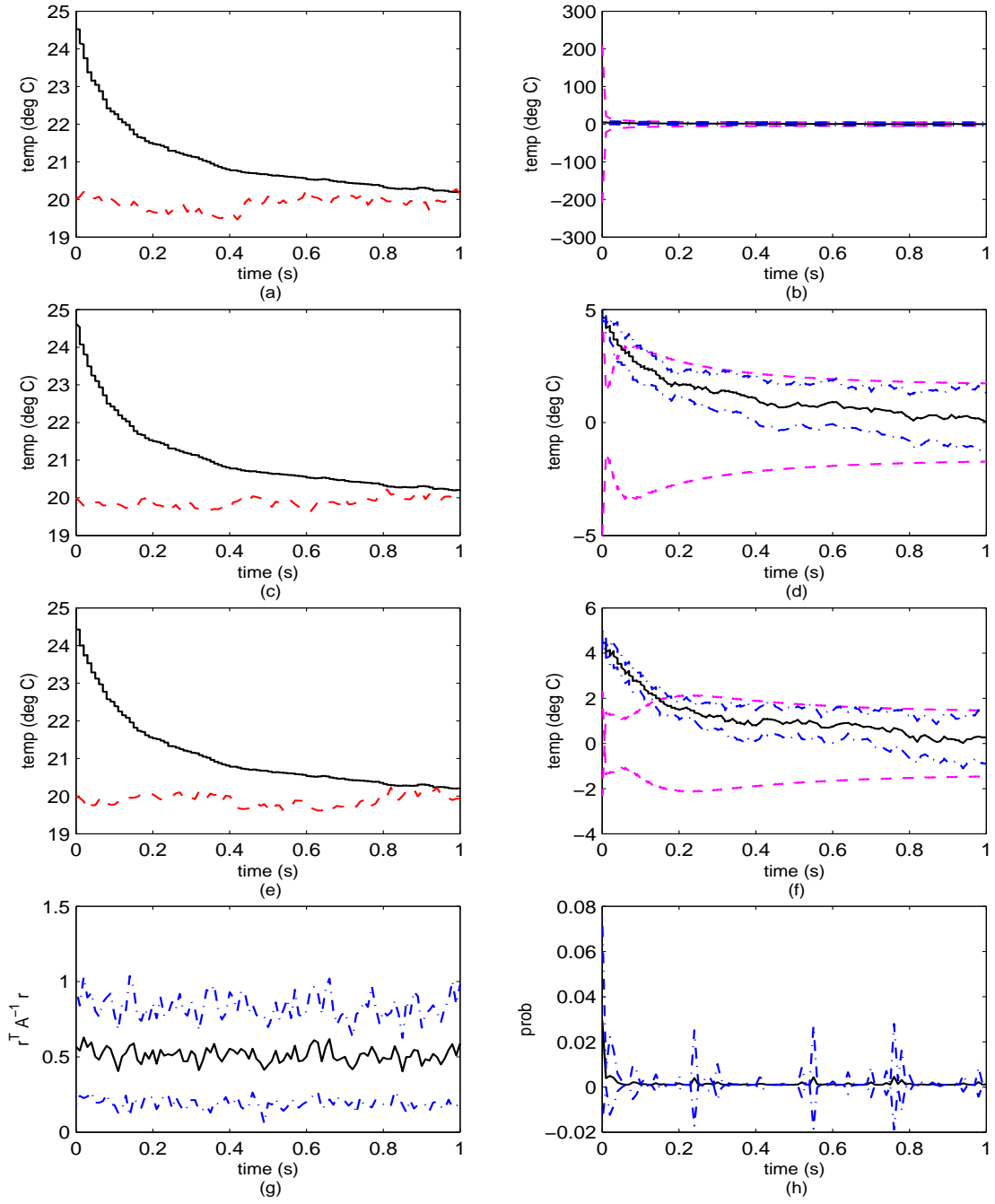


Figure 5.41 Simulation 3, Case 2: Elemental Filter 4. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

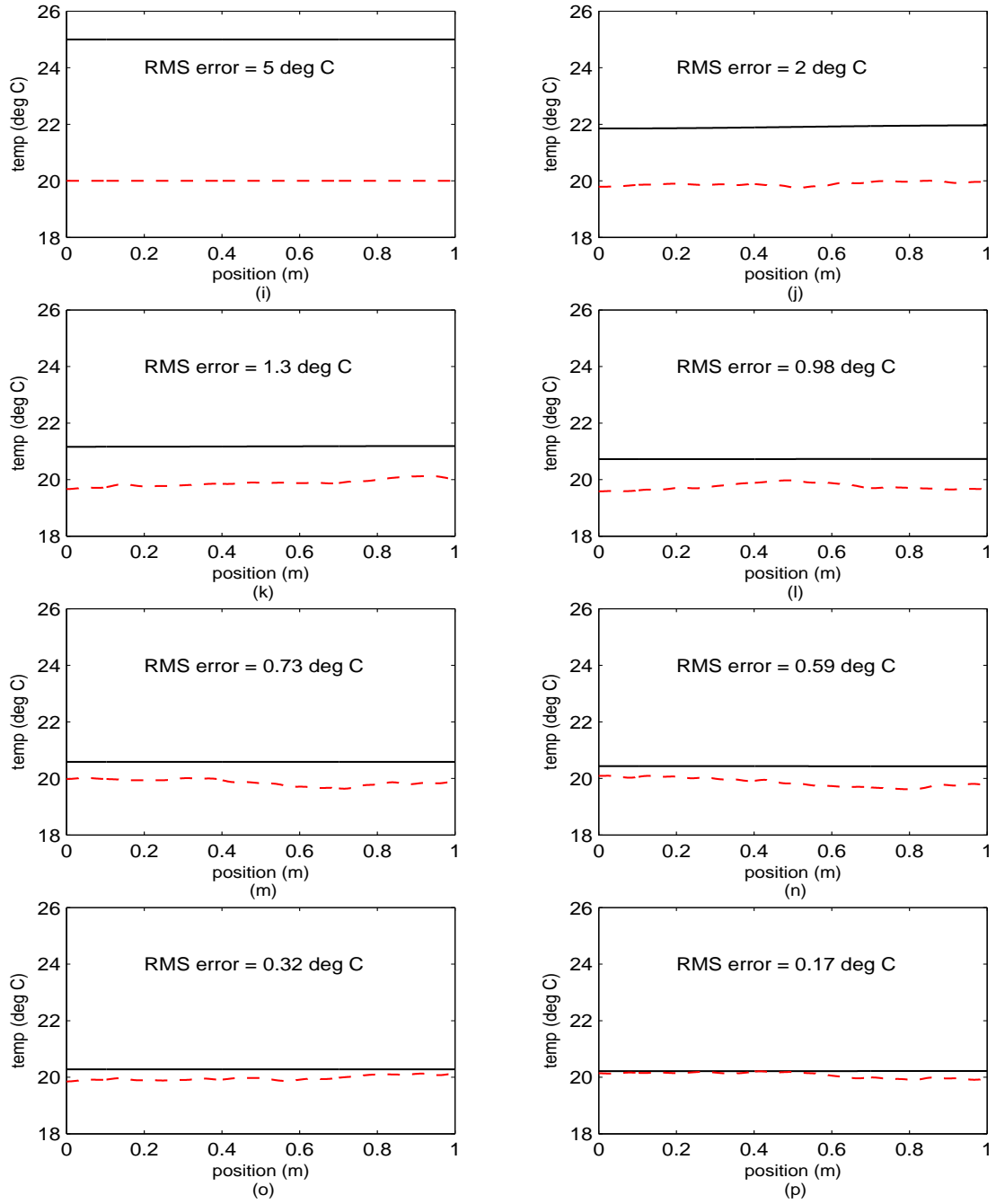


Figure 5.41 Simulation 3, Case 2: Elemental Filter 4 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

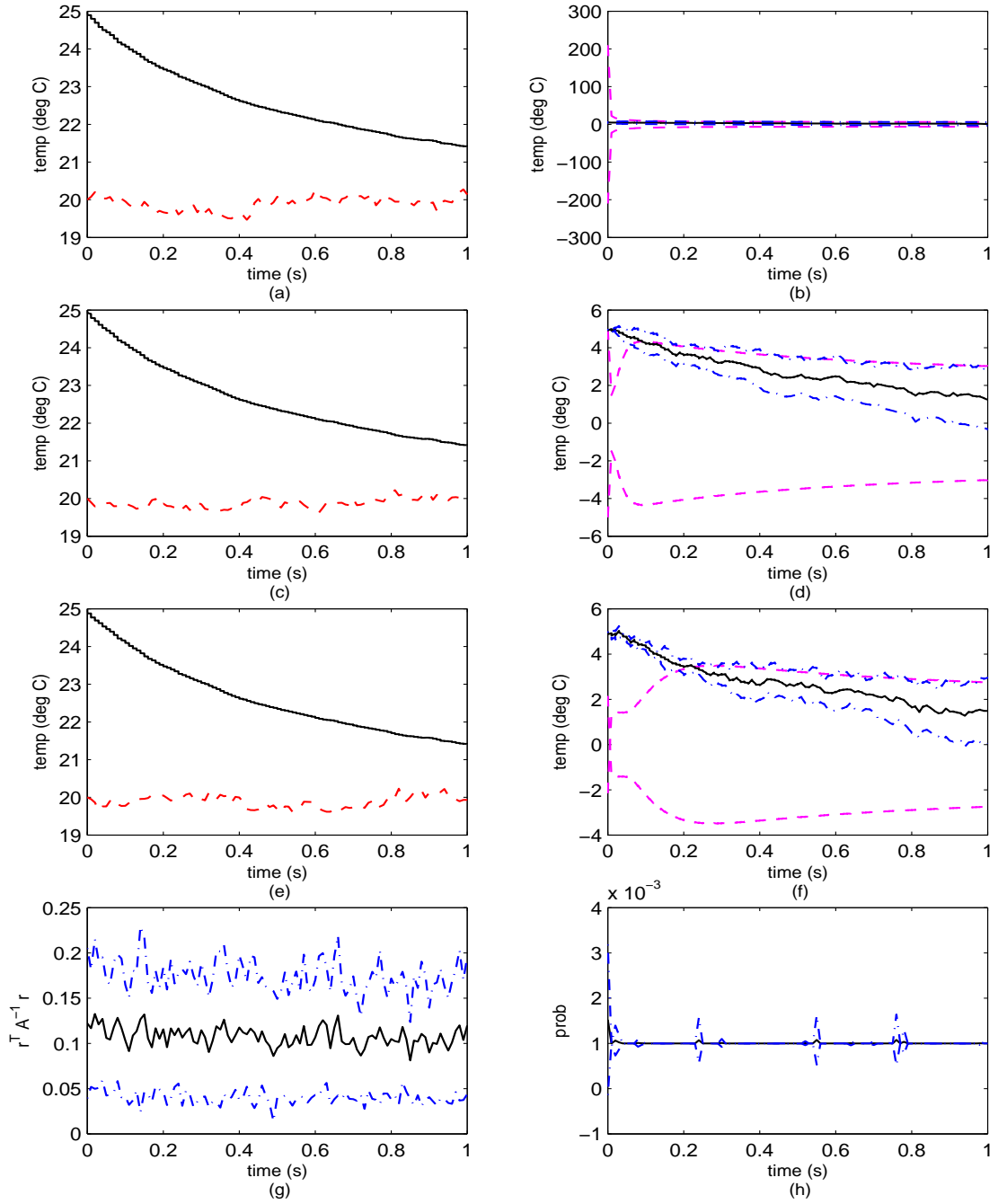


Figure 5.42 Simulation 3, Case 2: Elemental Filter 5. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

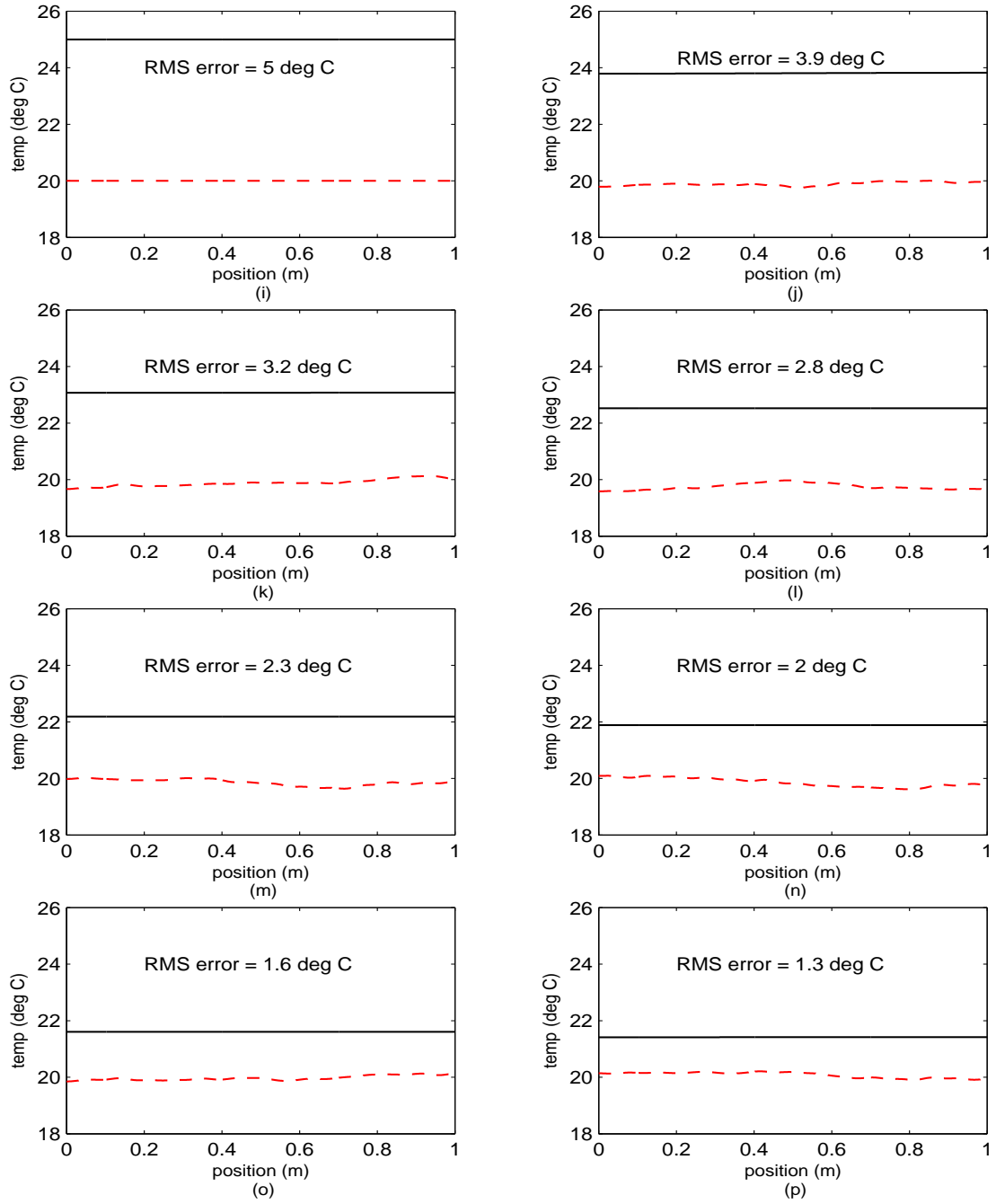


Figure 5.42 Simulation 3, Case 2: Elemental Filter 5 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

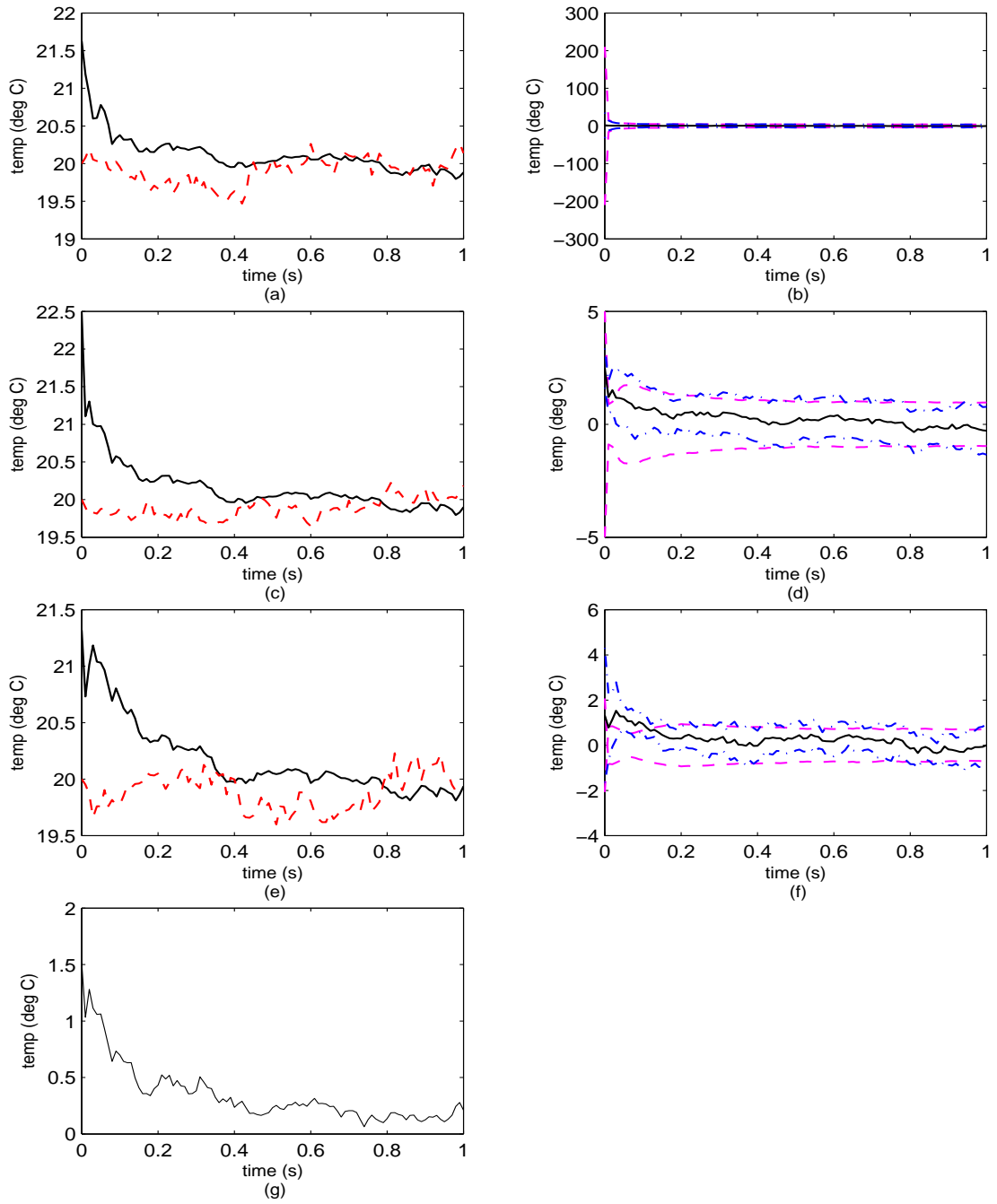


Figure 5.43 Simulation 3, Case 2: Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

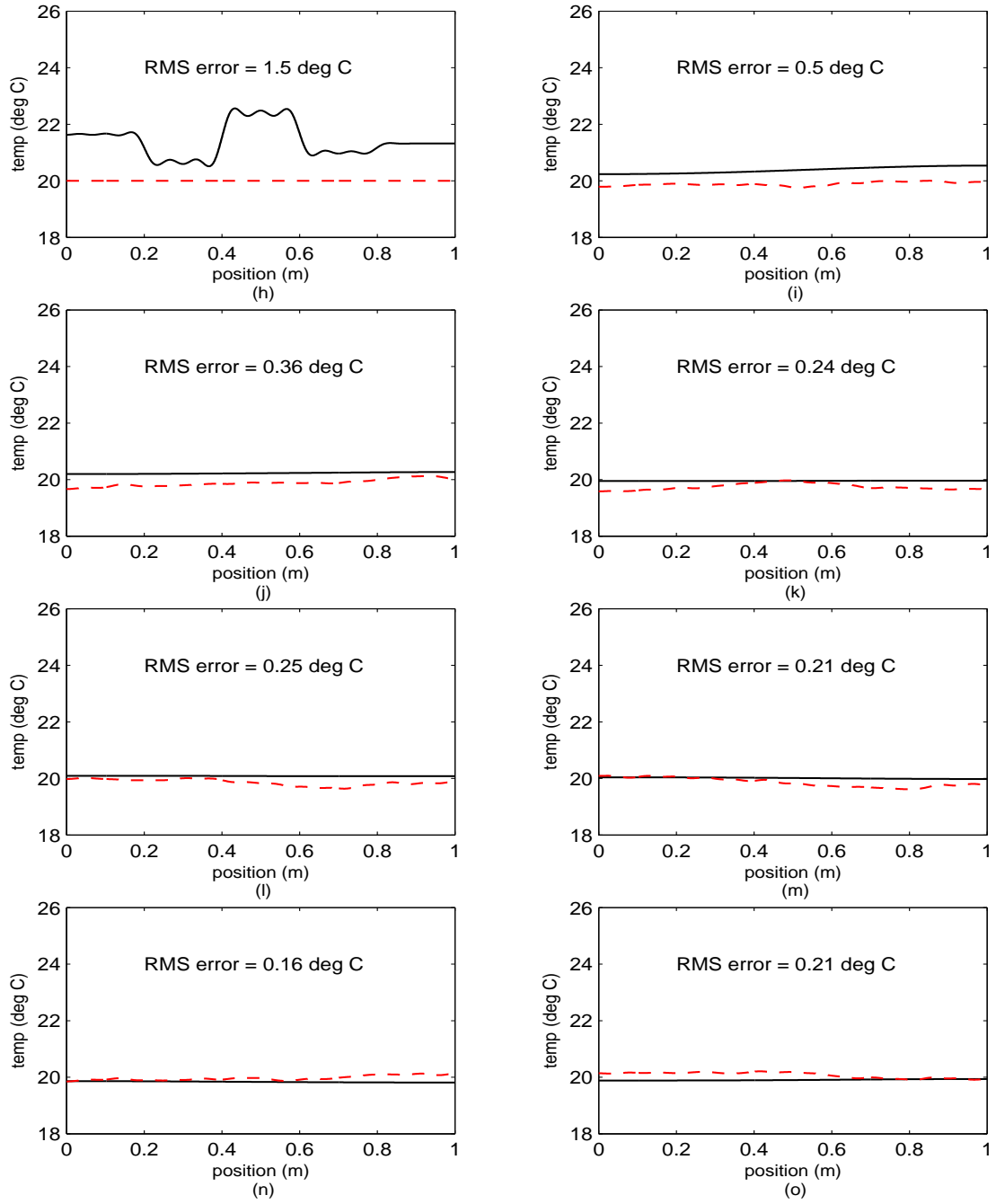


Figure 5.43 Simulation 3, Case 2: Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

Filter	Q	R	\hat{x}_0	P_0
1	Q_{true}	$0.04R_{\text{median}} = 0.4$	25	25
2		$0.2R_{\text{median}} = 2$		
3		$R_{\text{median}} = 10$		
4		$5R_{\text{median}} = 50$		
5		$25R_{\text{median}} = 250$		

Table 5.8 Simulation 4: Elemental Filter Parameters

5.5 Simulation 4

The preceding simulation showed how we could obtain a good estimate of an unknown *constant* R_{true} . This simulation demonstrates the capability of the MMAE to adapt to a time-varying R_{true} which varies over the interval $[R_{\text{min}}, R_{\text{max}}] = [1, 101]$. Specifically, we seek an accurate estimate of R_{true} as it either linearly increases or decreases during the one-second interval of interest. We shall use a bank of five elemental filters similar to those used in Simulation 3 — the difference lies in the R value used to center the filter bank. In the previous simulation, the center elemental filter was built using $R = 5$, whereas in this simulation, the median filter is “located” at $R = 10$ so that it coincides with the geometric mean of the minimum and maximum values for R_{true} . Thus, $R_{\text{median}} = \sqrt{R_{\text{min}}R_{\text{max}}} = \sqrt{101} \approx 10$. The relative spacing of the elemental filters is the same factor of five spacing, as are the rest of the elemental filter parameters. The elemental filter design parameters are tabulated in Table 5.8 for convenience.

We shall consider two distinct cases of linearly changing R_{true} : increasing and decreasing. The initial results for these cases are shown in Figures 5.44 and 5.52, respectively. The MMAE state estimate is quite good for both cases; the MMAE state estimate is off by less than 1 °C (RMS), as seen in Figures 5.51(o) and 5.59(o). Additionally, by inspecting Figures 5.44 and 5.52, we see that the probability flow among the elemental filters is slightly different, i.e., the figures are not mirror images of each other as we would expect if the MMAE handled R_{true} increases the same as

decreases. An increase in R_{true} creates a harsher noise environment for MMAE adaptation, versus a decreasing R_{true} , which presents a more benign noise environment. “Harsh” is in the sense that the measurements are less precise; “benign” is in the sense that the measurements are more precise.

Since the MMAE “prefers” a filter which overestimates the true measurement-corruption noise covariance to one which underestimates the true covariance, we should anticipate that the MMAE “switches” filters more quickly for the increasing true covariance case. Said another way, increases in the true covariance result in a more aggressive flow of probability to an elemental filter based on a larger covariance. For the decreasing true covariance case, an elemental filter based on a too-large value is not so quickly cast aside for a filter based on a more realistic model.

5.5.1 Simulation 4, Case 1 (Increasing R_{true}). For this first case, the true measurement-corruption noise covariance, R_{true} , varies linearly for times $t_i = \{0, 0.01, 0.02, \dots, 1\}$ according to

$$R_{\text{true}}(t_i) = 100(t_i + 0.01) \quad (5.12)$$

Thus R_{true} begins at 1 and ends at 101; this is shown graphically by the dashed line in Figure 5.44, as seen by noting the scale on the *right-hand side* of the plot. Note that the elemental filters in the bank completely cover this range of values for R_{true} and that no attempt has been made to optimize the placement of these filters.

At the beginning of this simulation, $R_{\text{true}}(t_1 = 0) = 1$ and elemental filters 1 and 2 represent the best hypotheses since they have R values of 0.4 and 2, respectively; see Figure 5.44. As the true R increases to 2 at the next time step, elemental filter 1 does not match quite as well and elemental filter 2 matches perfectly, however, it takes a few more update cycles for the change to be completely “noticed” and during that time, the true value has continued to increase and thus the next elemental filter (number 3) begins to become a better fit. Soon after elemental fil-

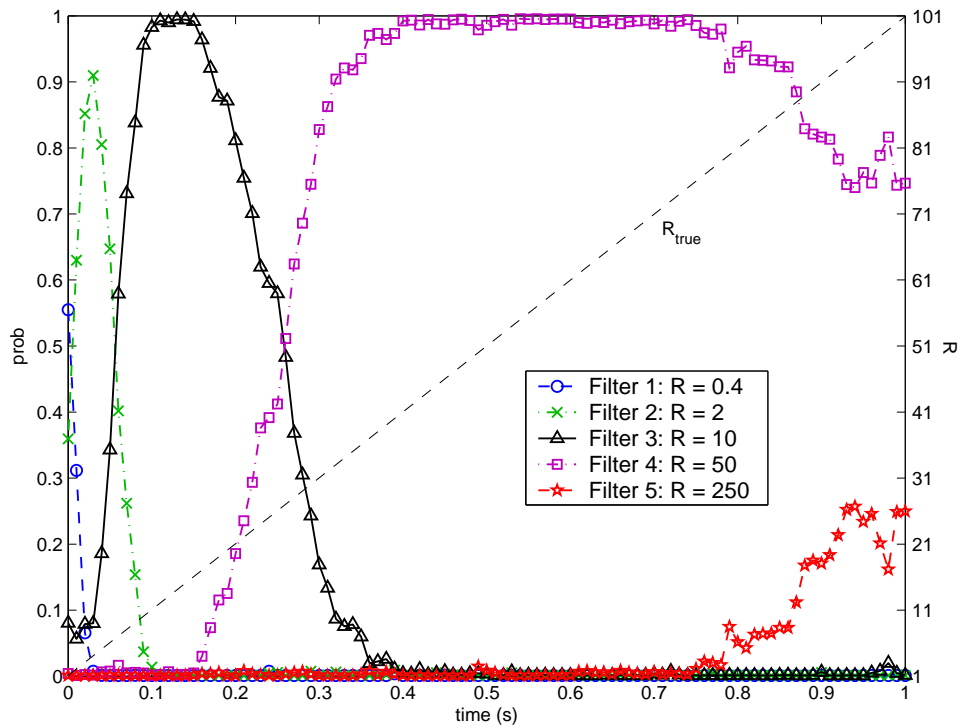


Figure 5.44 Simulation 4, Case 1 (increasing R_{true}): Hypothesis conditional probability.

ter 3 absorbs all of the probability, it too becomes less likely as the true R starts to “look” like the next filter in the bank, $R_4 = 50$. Note that only two elemental filters register an appreciable amount of probability at any one time, thus the usual situation is one in which an elemental filter either has a very good hypothesis, or the true value falls between the hypothesized values for neighboring filters.

The hypothesis conditional probability curves for each elemental filter resemble trapezoids which begin when the truth R is about one half of the filter-hypothesized R and that elemental filter remains the most likely elemental filter until the halfway point for the next elemental filter in the bank, as can be seen clearly in Figure 5.44. Therefore, we can predict that elemental filter 5 would become the most likely filter at time 1.25 seconds if we were to run the simulation for that length of time.

The second column in Table 5.9 gives the range of R_{true} values for the best filter. Note that the ranges are mutually exclusive. In the fourth column, note that

Filter	R	R_{true} when $p_k > 0.5$	R_{true} when $p_k > 0.1$
1	0.4	1	1, 2
2	2	2, ..., 6	1, ..., 9
3	10	7, ..., 26	5, ..., 32
4	50	27, ..., 101	19, ..., 101
5	250	—	88, ..., 101

Table 5.9 Simulation 4, Case 1 (increasing R_{true}): Best hypothesis. Filter 5 was never the best filter at any time during this simulation; hence the bar.

only two filters are “in force” at any one time with 10% or more likelihood. We could add yet another column corresponding to the R_{true} values when the standard deviation in the mean hypothesis conditional probability history is about zero; see plot (h) of Figures 5.46 through 5.50 at the end of this section. This represents the times when an elemental filter matched the true value without regard to the particular realization of measurement pseudonoise added. For example, elemental filter 3 is the best when $R_{\text{true}} = 10$, which is to be expected, up until $R_{\text{true}} = 14$, which is a fairly small window given the neighboring filter is at $R = 50$. Elemental filter 4 has a much wider “perfect” match zone, from $R_{\text{true}} = 40$ until $R_{\text{true}} = 64$. Furthermore, the first few sample periods embody the usual initial transient in each of the elemental filters and thus these first few time instants are not particularly indicative of the true capabilities of the MMAE. For instance, in this research, all of the elemental filters in the bank are assumed to be equally likely when the simulation begins; thus, even the most mismatched filter gets $1/K$ th of the probability at the start.

An interesting trend that we can easily see in the plot (g) likelihood quotient histories is readily explained using the formula developed in Section 5.1 for the steady state likelihood quotient for the k th elemental filter at time t_i (repeated here for our convenience)

$$E\{\mathbf{L}_k(t_i)|_{t_i=t_{\text{ss}}}\} \cong \frac{R_{\text{true}}}{R_k} M \quad (5.10)$$

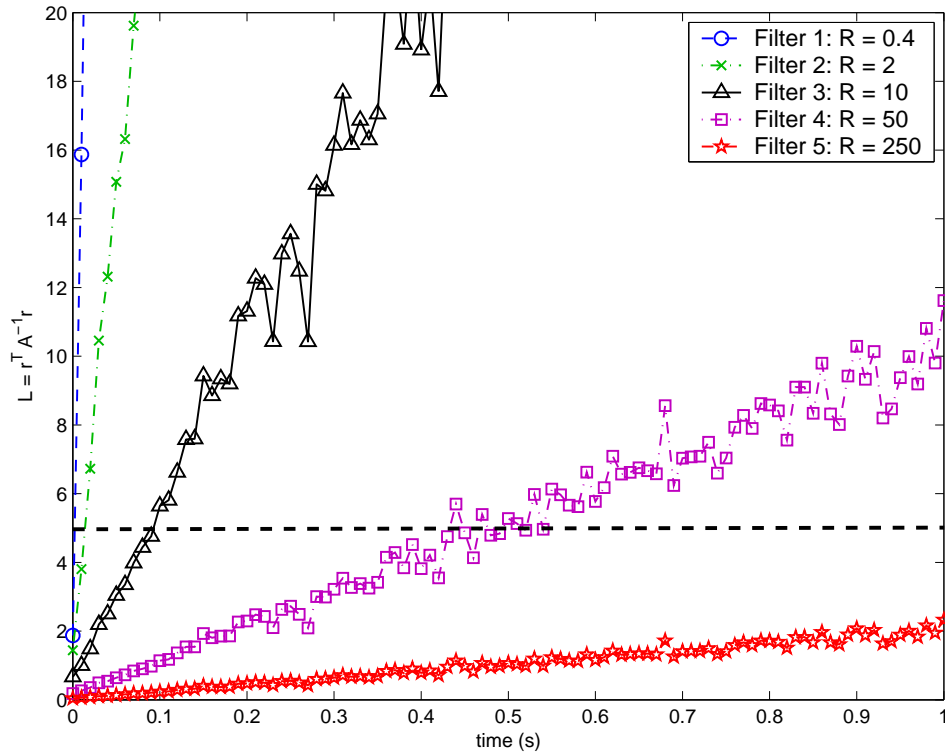


Figure 5.45 Simulation 4, Case 1 (increasing R_{true}): Likelihood quotient. This plot has been clipped at $L = 20$ since only the elemental filters operating near $L = 5$ absorb an appreciable amount of probability.

As the truth value for R increases, so do the likelihood quotients (which of course never really make it to steady state) since we see that R_{true} is in the numerator, while M/R_k remains unchanged. Thus, the R_{true} “ramp” is apparent in the likelihood quotient “ramps” as seen in Figure 5.45.

In general, we note that the filters with smaller assumed values for R gave better state estimates, because in the beginning when R_{true} was also small, we had the case of precise measurements; hence estimation is very good. Specifically, elemental filter 1 converged rapidly because it accurately reflected the high quality of the measurements; see Figure 5.46. Recall that when R is small, the (Q/R) ratio is large; therefore the gain is large and that gives us the rapid convergence. However, as the measurement quality waned, the state estimate started to “wander” and its RMS error increased as shown in Figures 5.46 (i) through (p). At the other end of

the filter bank, elemental filter 5 showed a relatively large RMS error [Figure 5.50 (i) through (p)] because it initially assumed that the measurements were “rough”, and then as time progressed, the measurements truly did fall in quality.

Overall, the MMAE state estimate is quite good given the fact that we rarely have an elemental filter based on the perfect model. Because the elemental filters are more closely spaced at small values of R , the RMS state error is smaller at the beginning (after the initial transient) of the simulation as seen in Figure 5.51(g) for all time and then in plots (h) through (o) (on the following page) along the length of the rod. Of course, the fact that a small truth R simply means that the measurements are more precise contributes to our excellent performance for small R_{true} . Since we knew a priori that R was going to change linearly, perhaps we could have improved the state estimation performance by spacing the elemental filters linearly. An inspection of Figure 5.45 shows that the likelihood quotient histories are all roughly linear. Thus, the quadratic nature of the likelihood quotient, $L_k(t_i) = \mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)$, with respect to the residuals matches up well with a linear change in the measurement-corruption noise covariance and thus a strictly linear spacing of the elemental filters would have given an unnecessary concentration of filters at larger values of R .

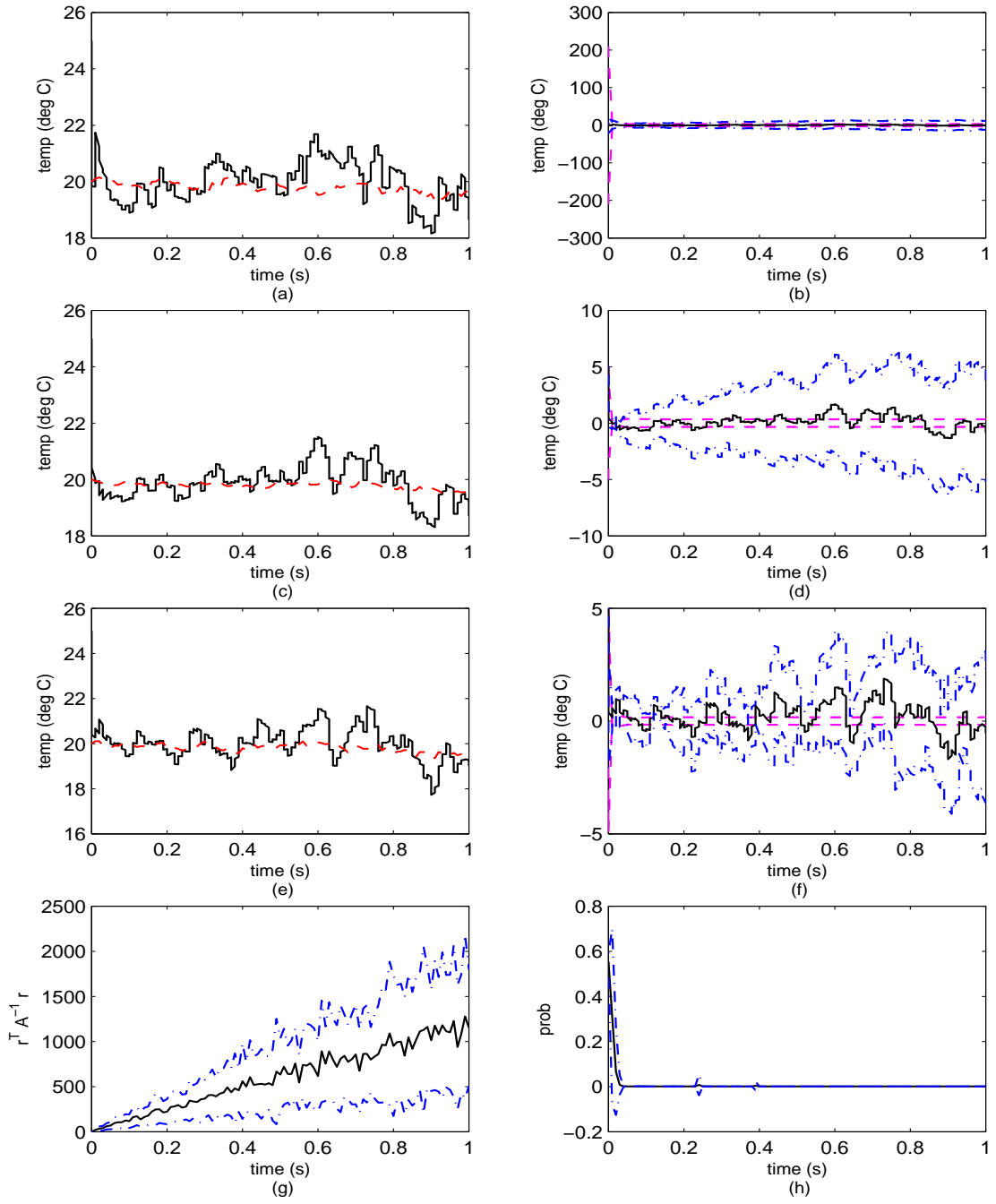


Figure 5.46 Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

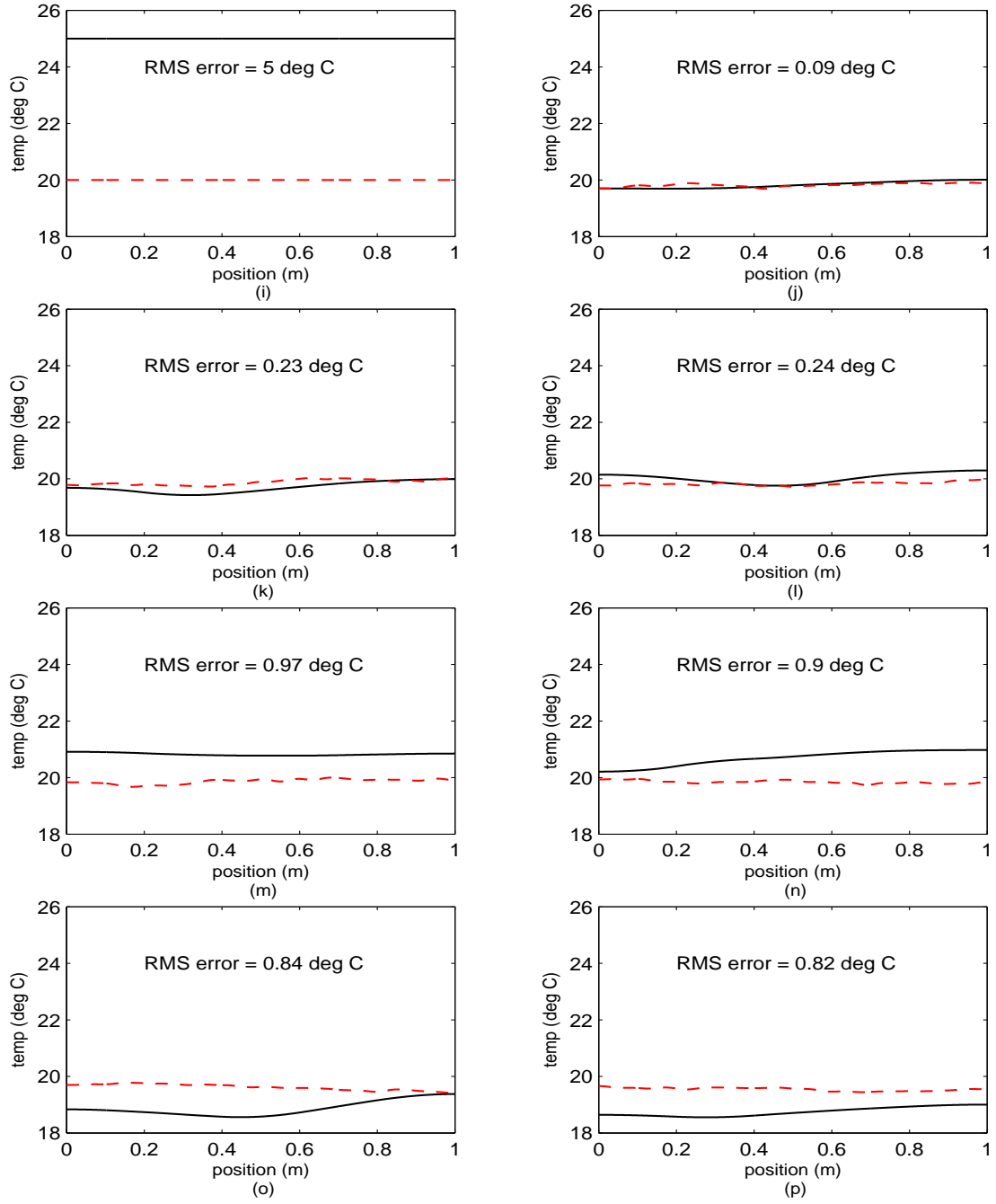


Figure 5.46 Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

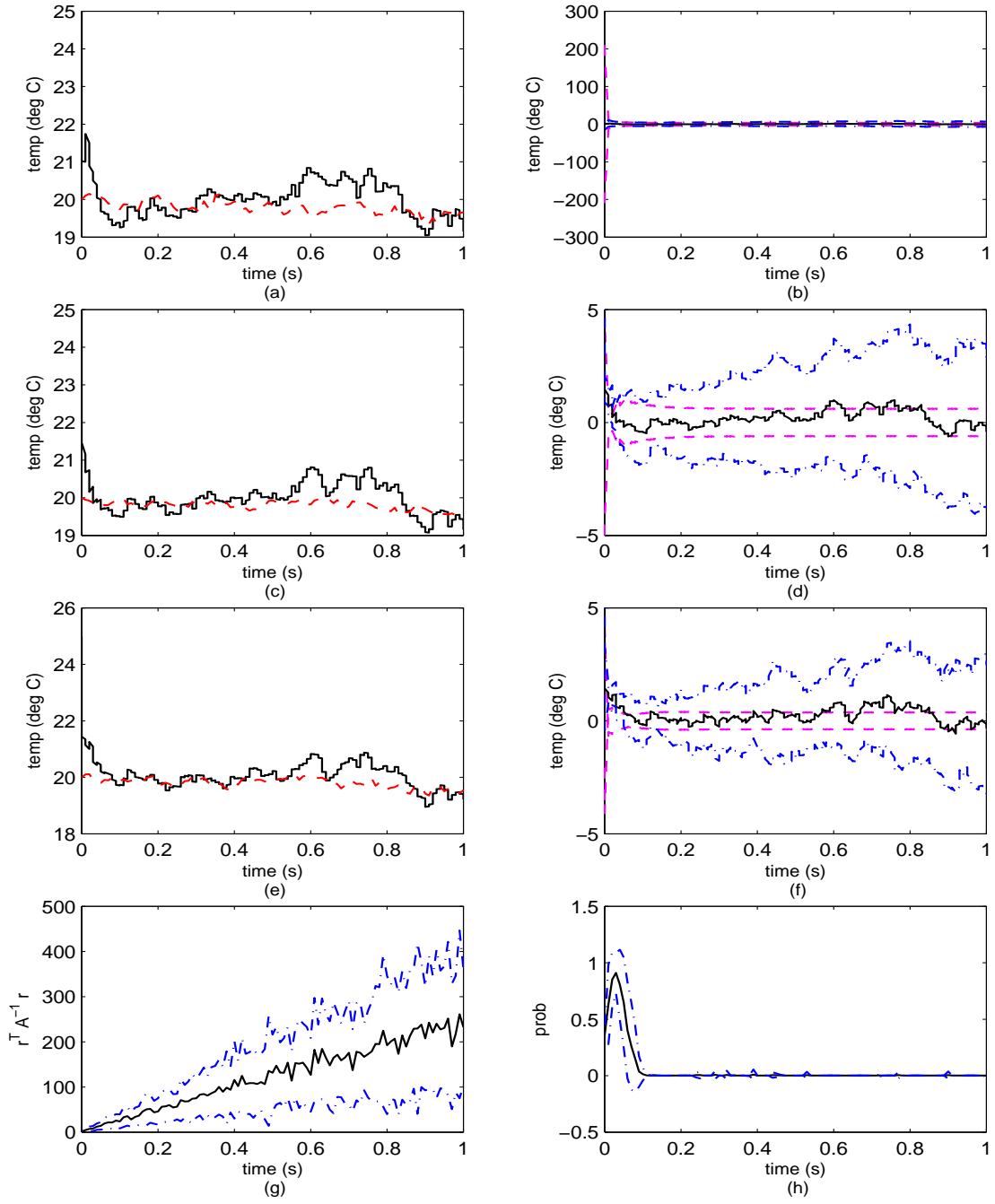


Figure 5.47 Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

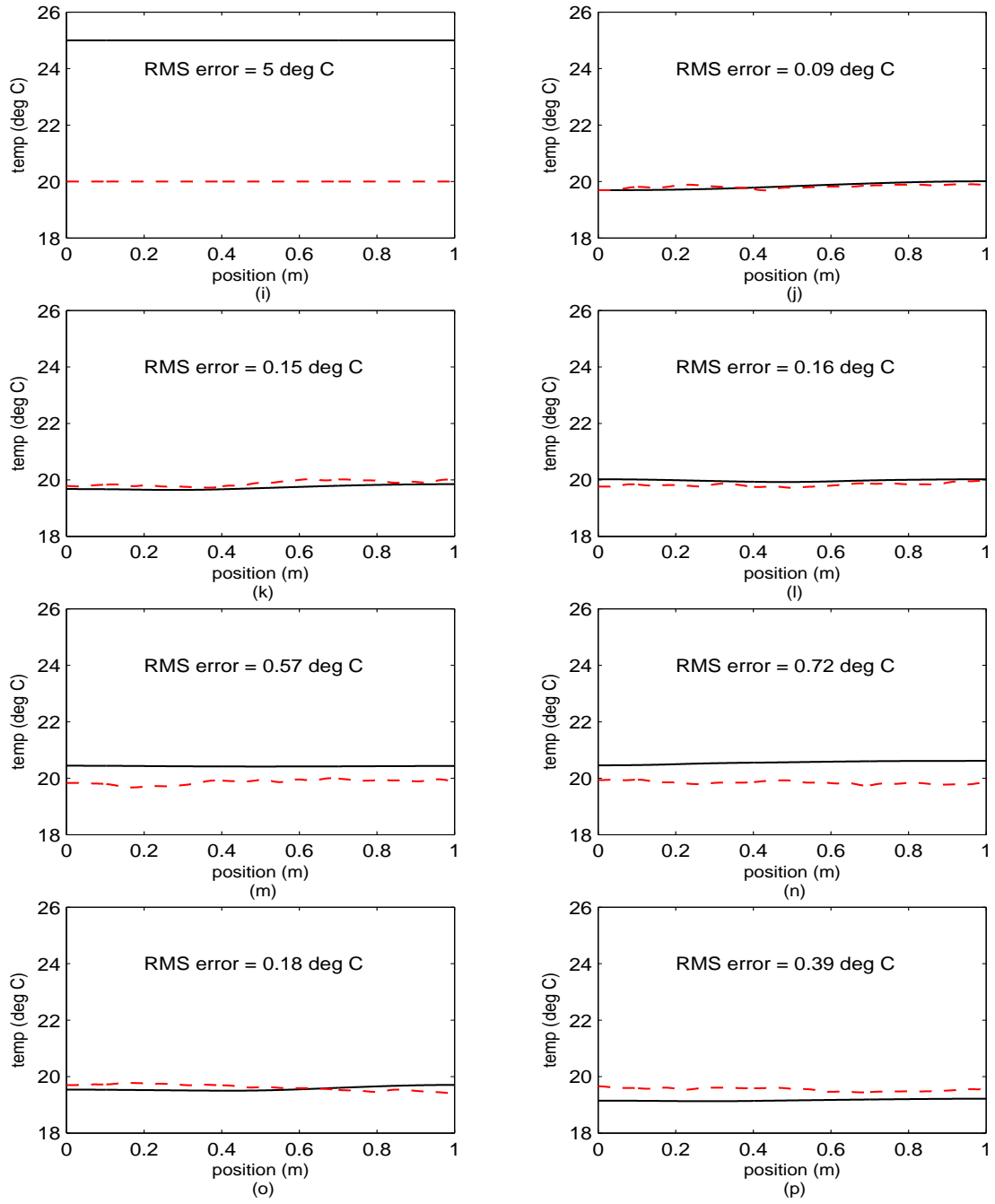


Figure 5.47 Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

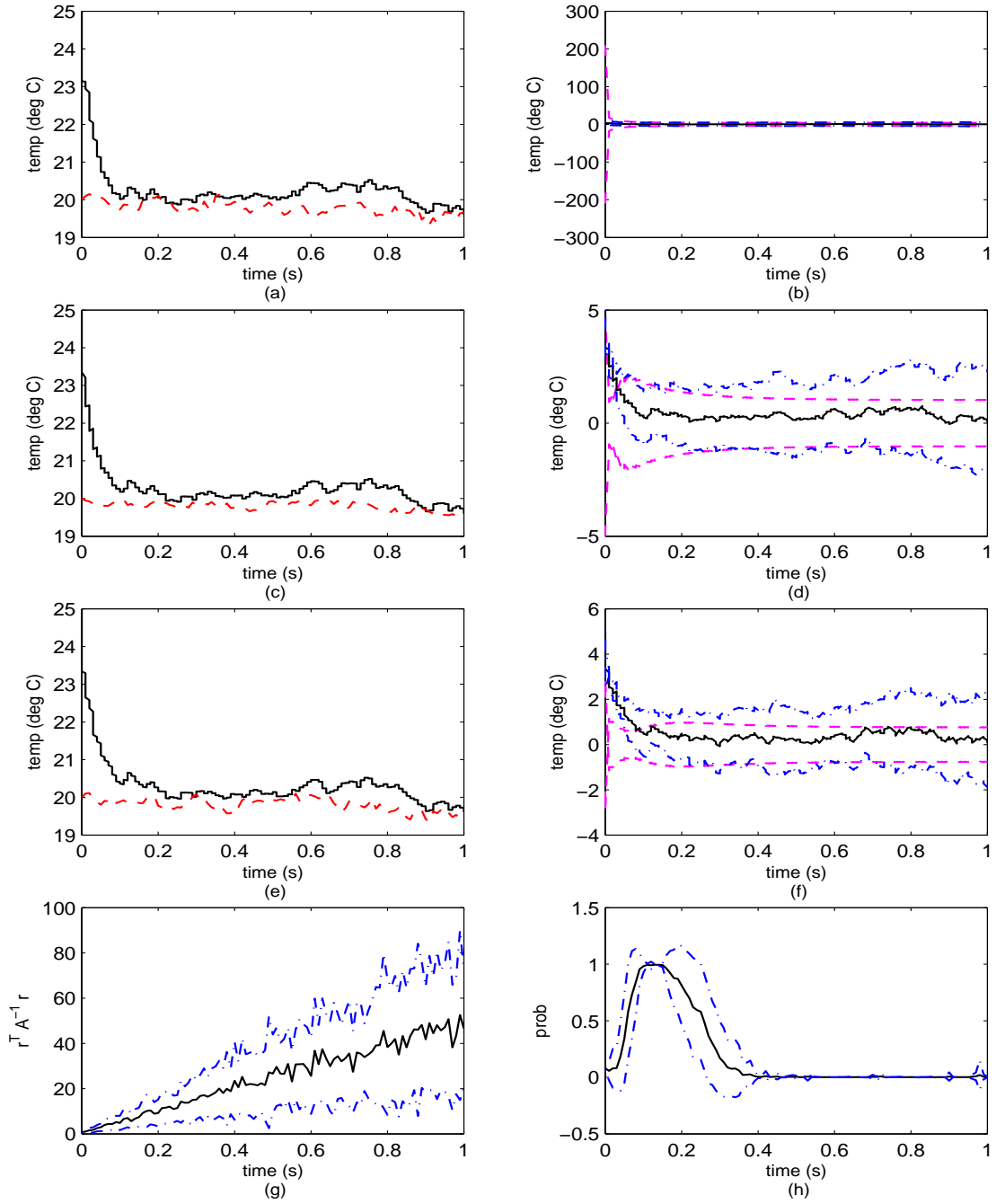


Figure 5.48 Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

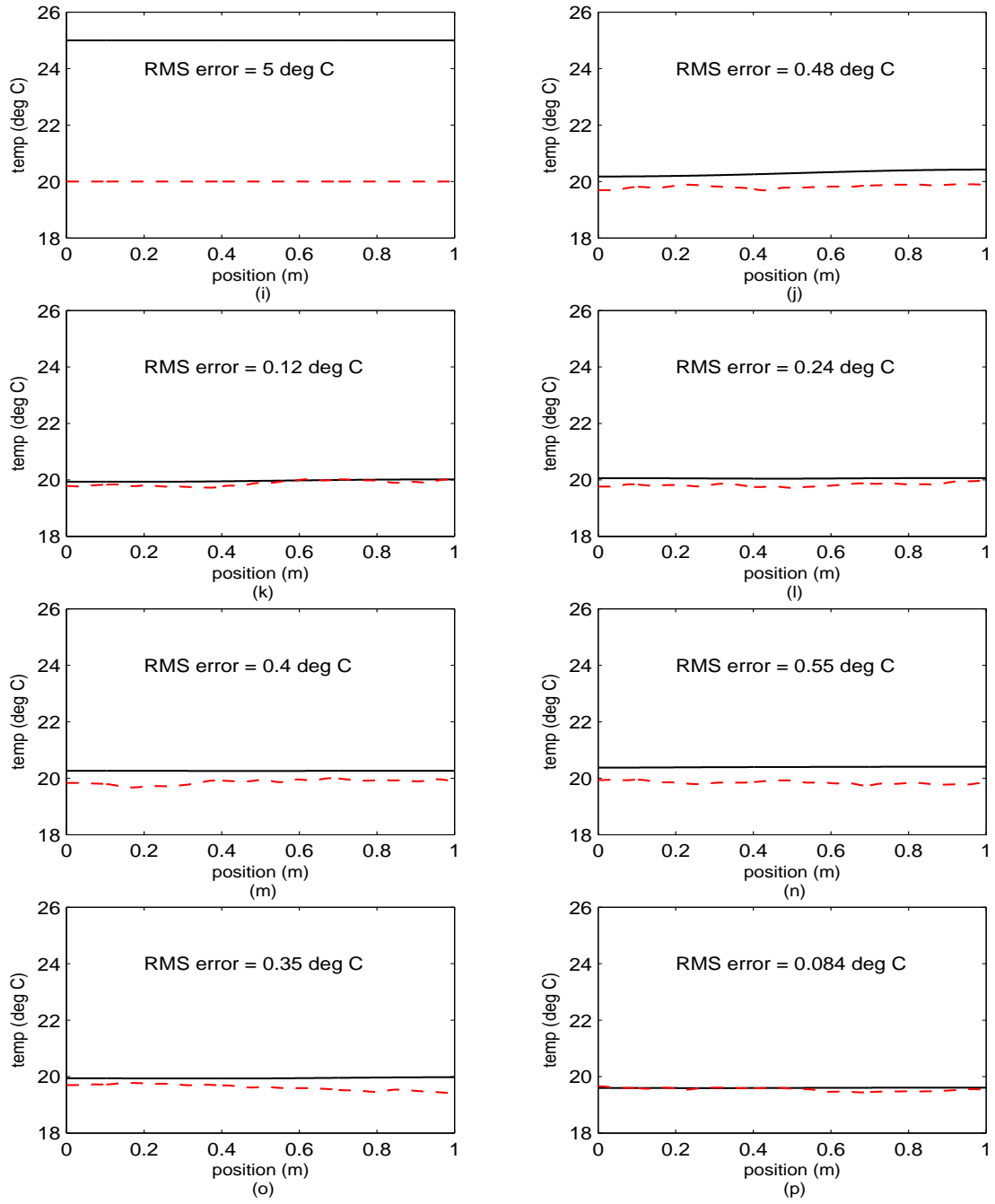


Figure 5.48 Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

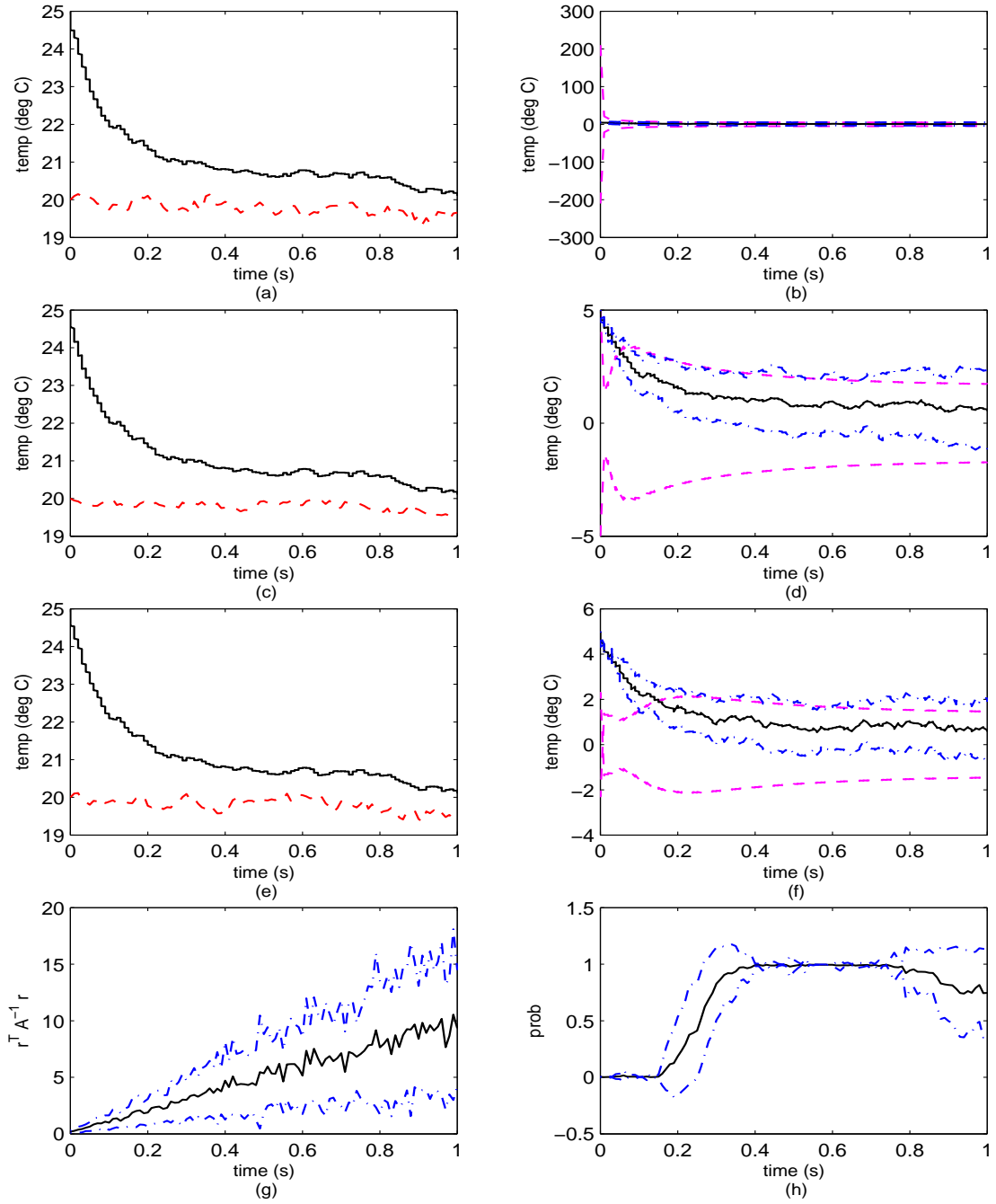


Figure 5.49 Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 4. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

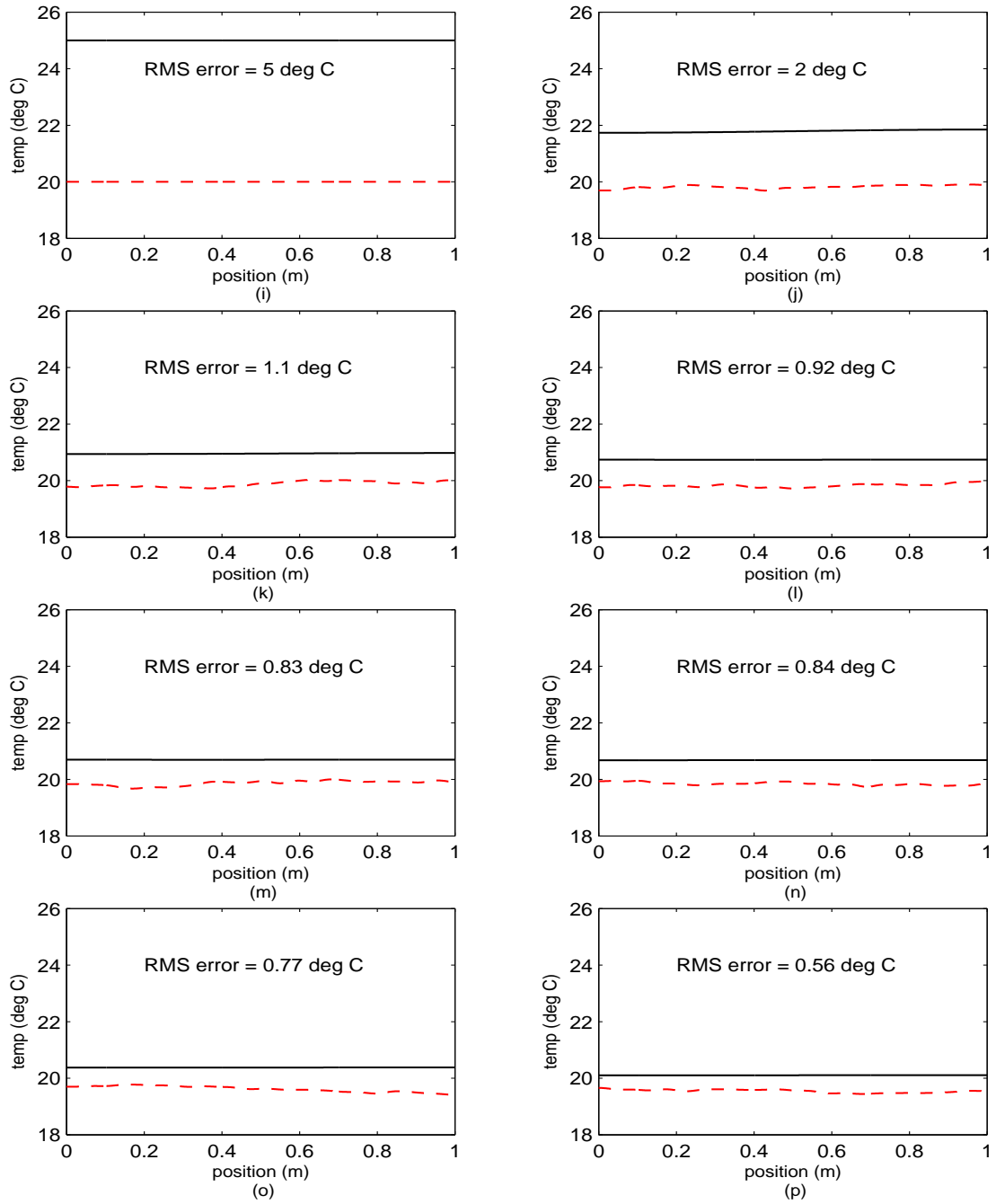


Figure 5.49 Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 4 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

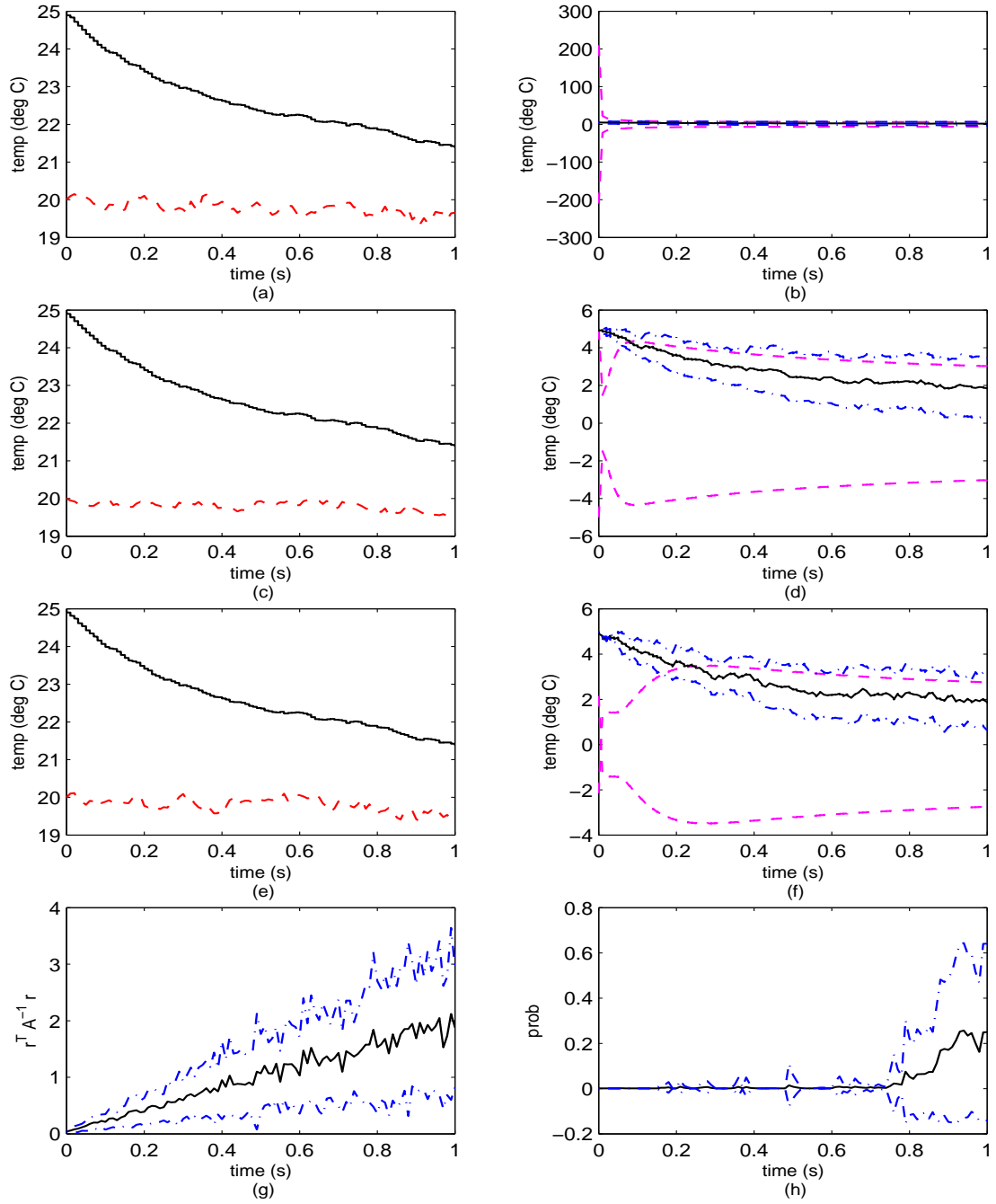


Figure 5.50 Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 5. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

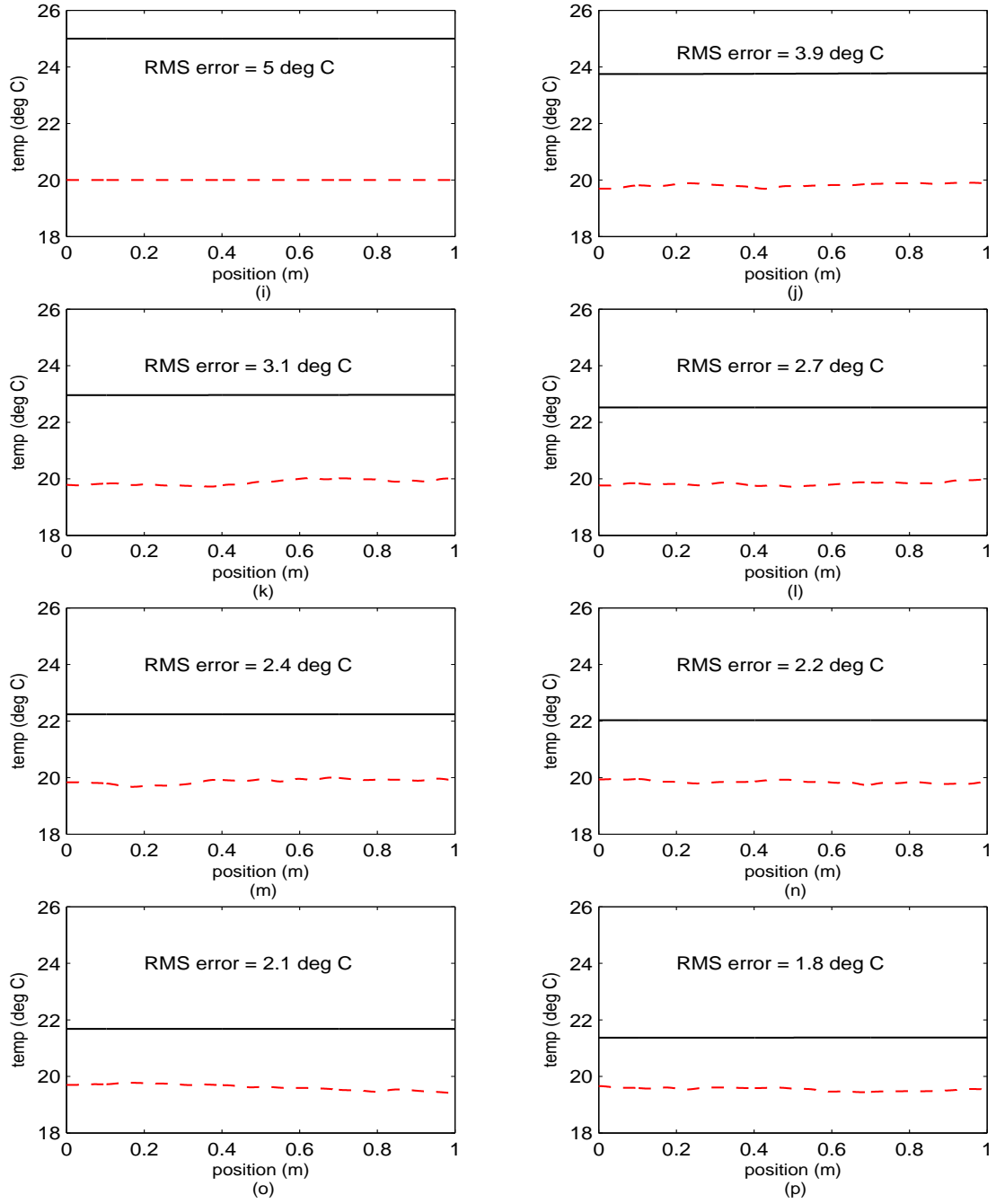


Figure 5.50 Simulation 4, Case 1 (increasing R_{true}): Elemental Filter 5 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

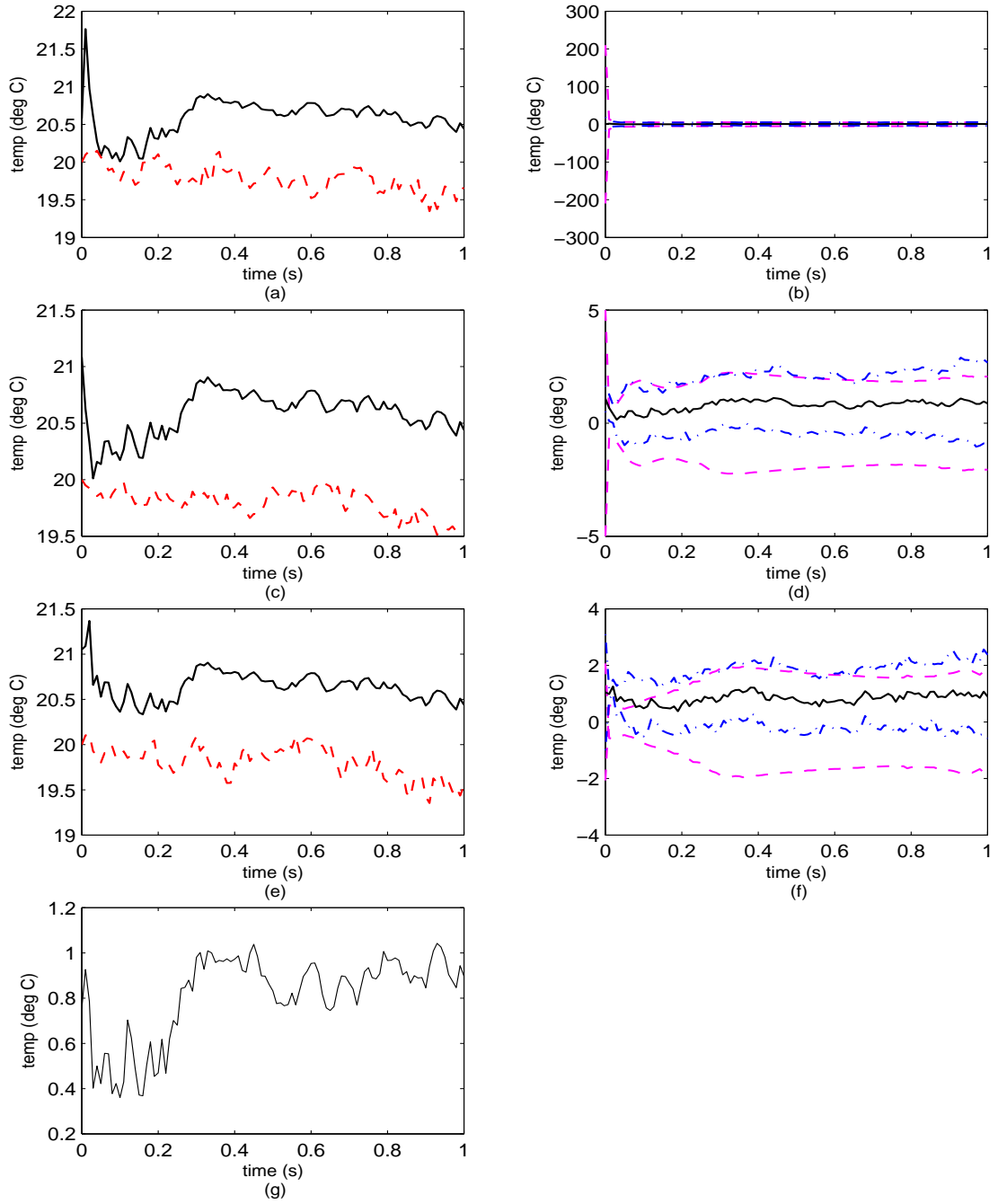


Figure 5.51 Simulation 4, Case 1 (increasing R_{true}): Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

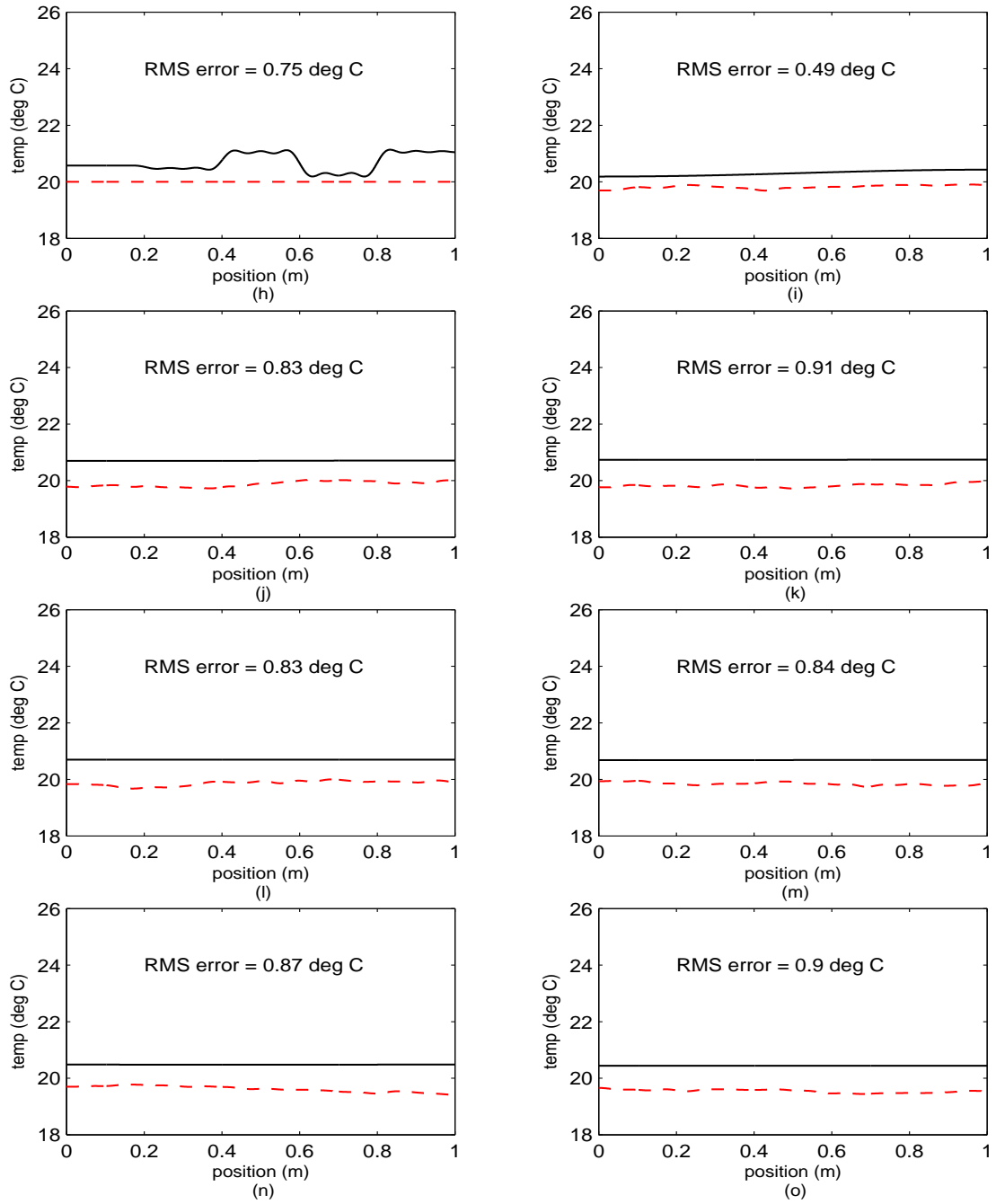


Figure 5.51 Simulation 4, Case 1 (increasing R_{true}): Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

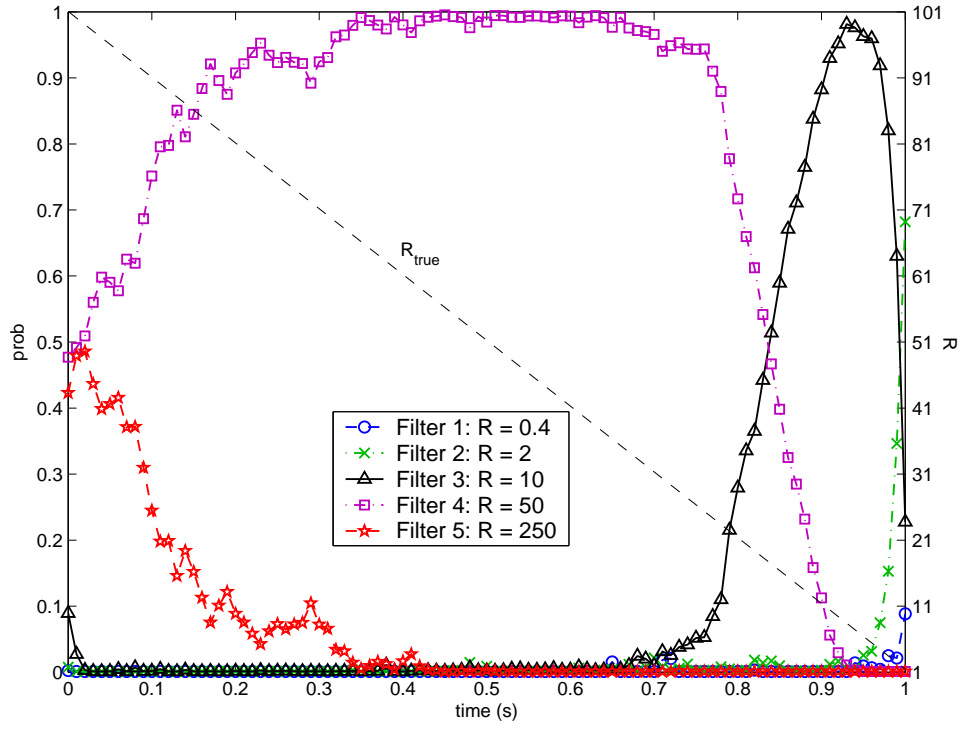


Figure 5.52 Simulation 4, Case 2 (decreasing R_{true}): Hypothesis conditional probability.

5.5.2 *Simulation 4, Case 2 (Decreasing R_{true}).* In this linearly decreasing case, the true measurement-corruption noise covariance eigenvalues were varied for $t_i = \{0, 0.01, 0.02, \dots, 1\}$ according to

$$R_{\text{true}}(t_i) = 100(1.01 - t_i) \quad (5.13)$$

Thus R_{true} begins at 101 and decreases to 1. Figure 5.52 shows the probability flow between the elemental filters as R_{true} changes.

The second column in Table 5.10 gives the range of R_{true} values for the best filter. Note that the ranges are mutually exclusive. In the fourth column, note that only two filters are “in force” at any one time with 10% or more likelihood. As we mentioned in Case 1, we could also note that, by looking at the (h) plots for the elemental filters, we could assess the times when the filter is almost perfect as

Filter	R	R_{true} when $p_k > 0.5$	R_{true} when $p_k > 0.1$
1	0.4	—	—
2	2	1	1, 2, 3
3	10	2, ..., 17	1, ..., 23
4	50	18, ..., 101	10, ..., 101
5	250	—	85, ..., 101

Table 5.10 Simulation 4, Case 2 (decreasing R_{true}): Best hypothesis. Filters 1 and 5 were never the best filters during this simulation; hence the bars.

indicated by the lack of variation in the hypothesis conditional probability curves. For example, elemental filter 4 shows very little variation from time 0.45 seconds until 0.65 seconds when R_{true} decreased from about 55 to 35. This is slightly different from the range encountered when R_{true} is increasing: that range is roughly 40 to 65. Thus for Case 2, we can see that the MMAE does not want to give up on the elemental filter that has overestimated the severity of the noise (elemental filter 5) so readily for the filter that is more conservatively modeled (elemental filter 4) while the true measurement-corruption noise covariance is decreasing. This phenomenon can be readily explained by using the likelihood quotient, $L_k(t_i) = \mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)$, values given for the two elemental filters of concern. In this decreasing R_{true} case, the MMAE seems to “hold” onto an elemental filter based on a too high assumption for the measurement-corruption noise covariance R . See Figure 5.53 for the likelihood quotient histories for all five elemental filters; note that they are all linear (with a negative slope) in response to the decreasing linear change in the measurement-corruption noise covariance — the filter in force is the filter with an L_k closest in value to $M = 5$. This tendency to hold on to an elemental filter based on a too high assumption for R can be explained by noting that the slope of the likelihood quotient history for the fourth elemental filter is greater than the slope for the fifth elemental filter; thus, while elemental filter 5 simply attempts to maintain its share of the probability by slowly diverging from a likelihood, L_5 , near five, elemental filter

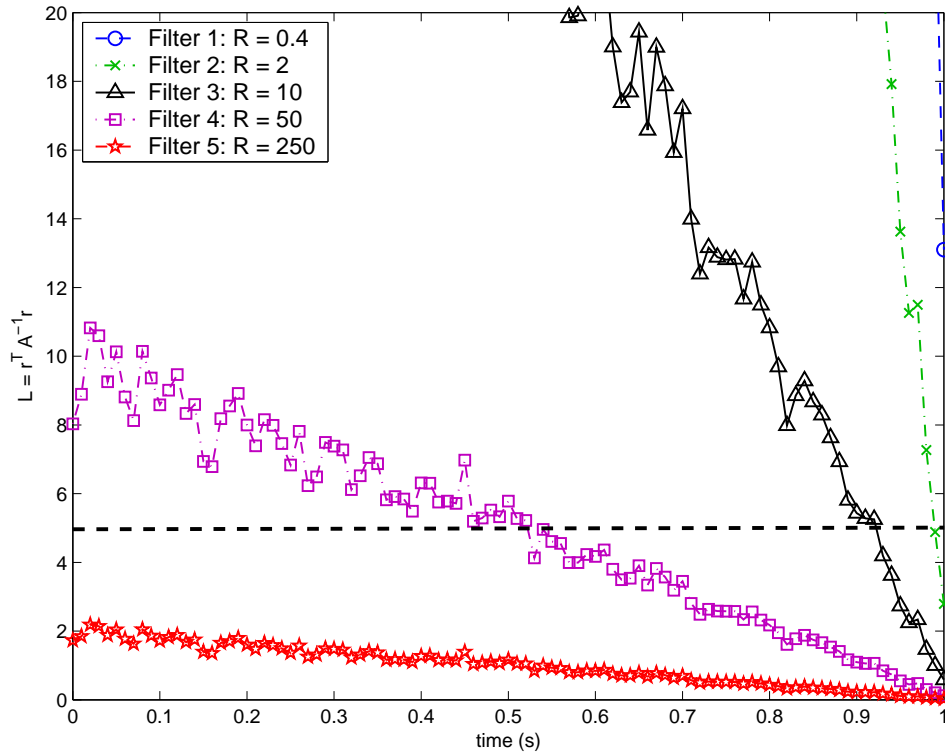


Figure 5.53 Simulation 4, Case 1 (decreasing R_{true}): Likelihood quotient. This plot has been clipped at $L = 20$ since only the elemental filters operating near $L = 5$ absorb an appreciable amount of probability.

4 works harder to gather all of the probability flow unto itself by rapidly approaching $L_4 = 5$.

Overall, the MMAE state estimate was better for the increasing measurement-corruption noise covariance case, although the final estimate at 1.0 seconds was better for this decreasing covariance case because we had nearly perfect measurements during the last 0.3 seconds of this simulation as seen in Figure 5.59 in plots (m) through (o). As stated earlier, we could have anticipated this overall result by noting that, as R_{true} decreases, the elemental filter in force with the bulk of the probability is still correct to a large degree, but gradually, the true error covariance of the measurements tightens up and an elemental filter with a smaller covariance is slowly promoted. By not adapting more quickly, the state estimate suffers some. However, while the initial state estimate bias was partially masked by the high R_{true} ,

the steadily improving measurements resulted in a steady convergence to the true temperature profile along the rod as seen in Figure 5.59 (g) for all time and then in plots (h) through (o) (on the following page) along the length of the rod.

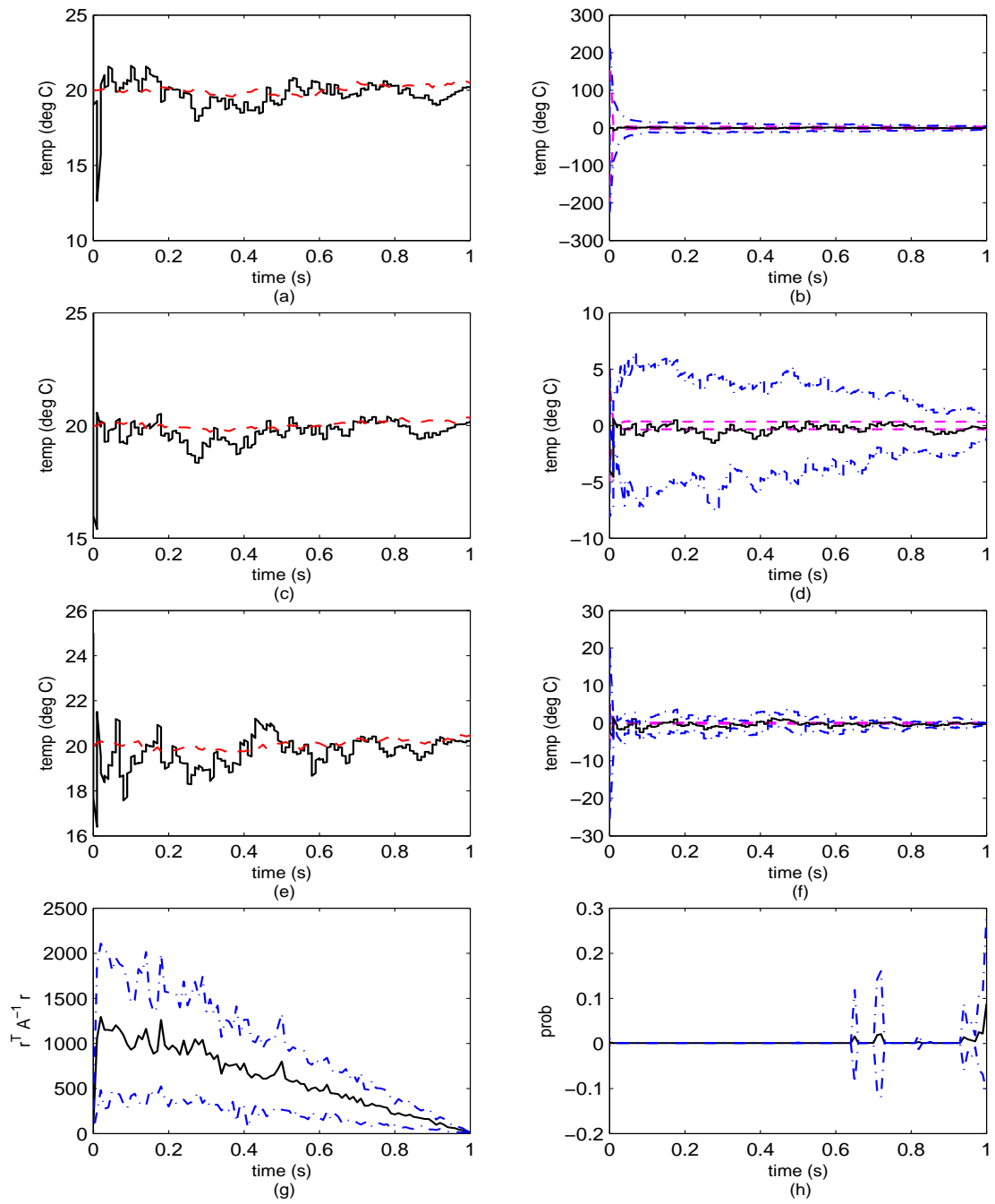


Figure 5.54 Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

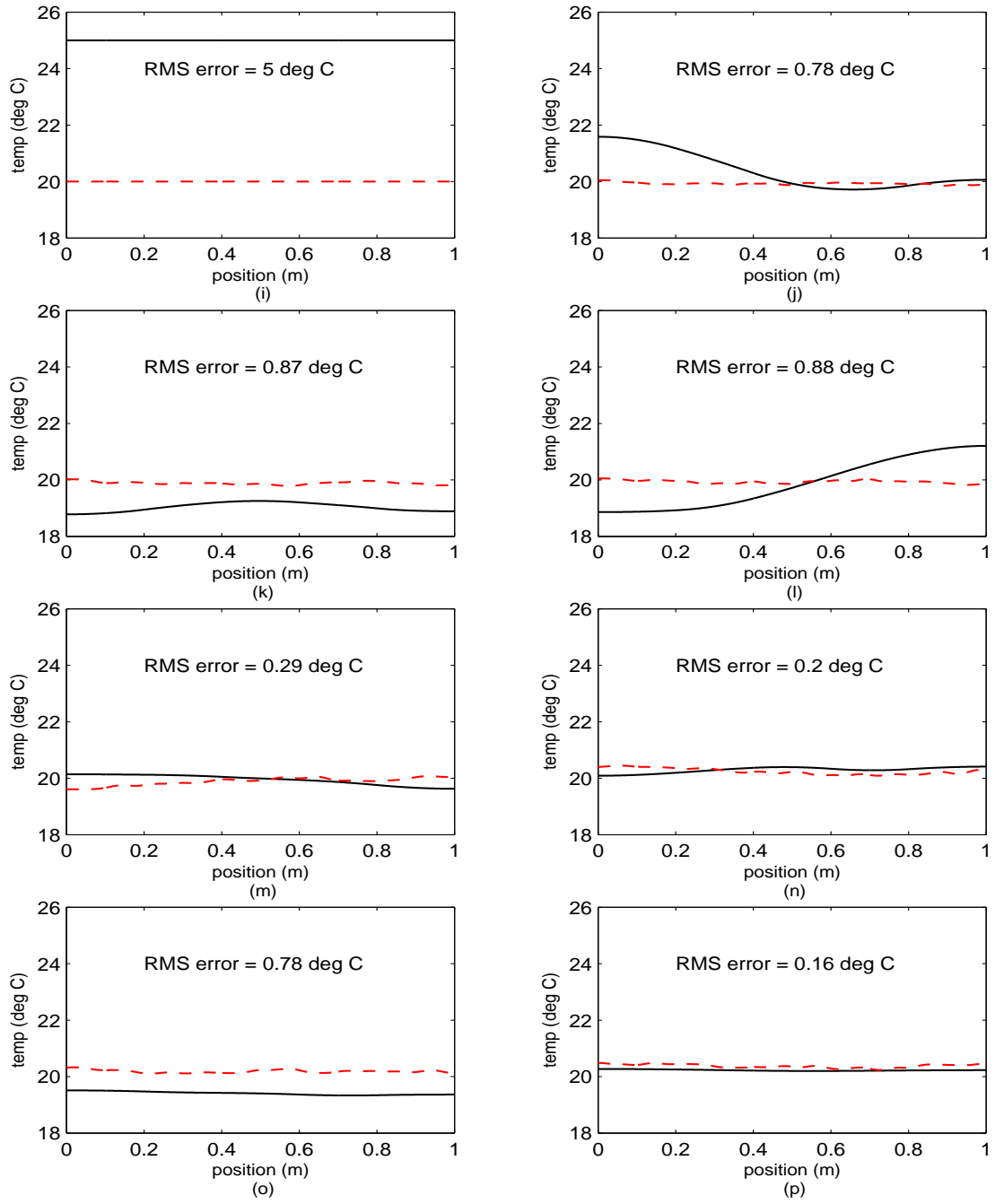


Figure 5.54 Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

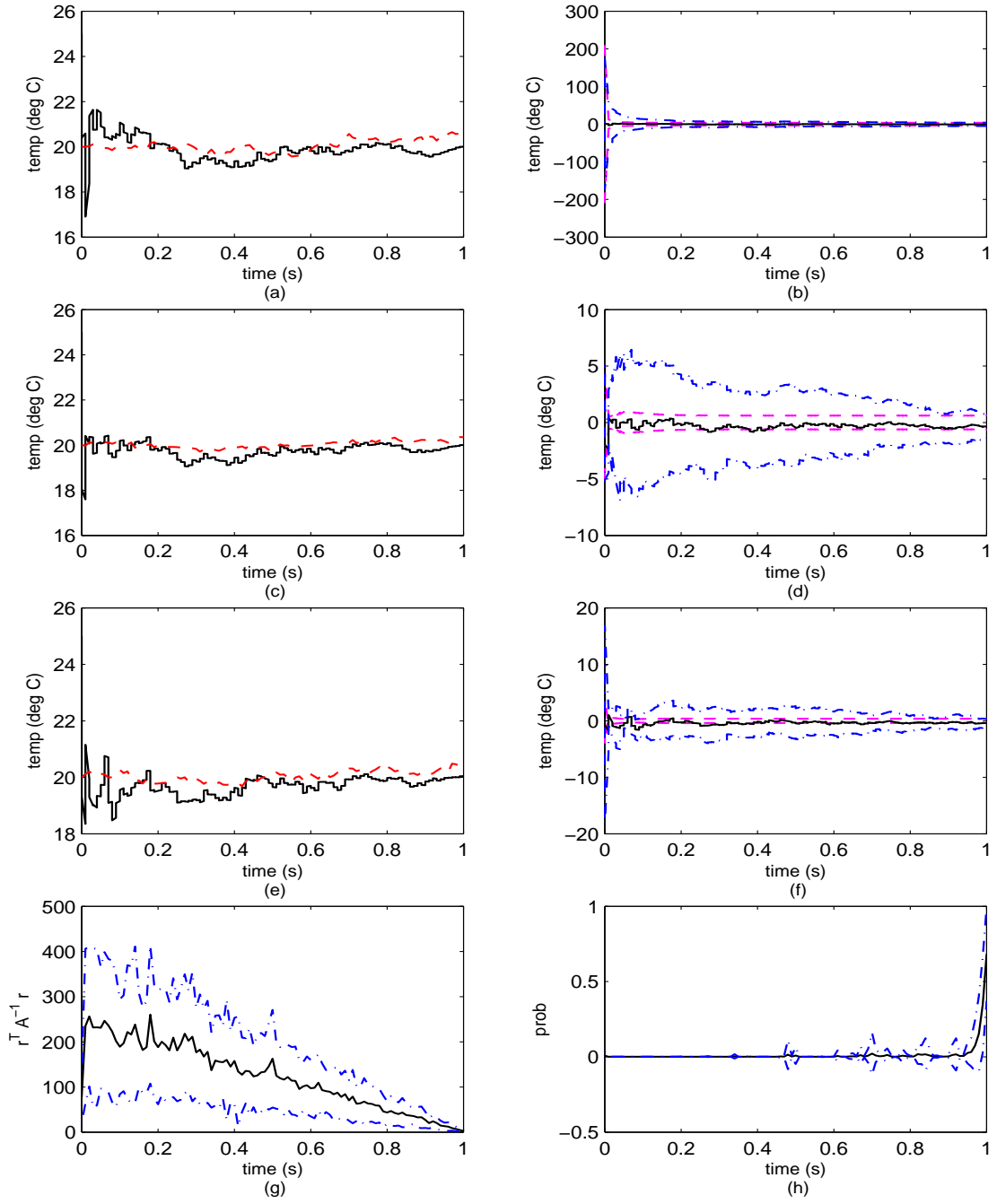


Figure 5.55 Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

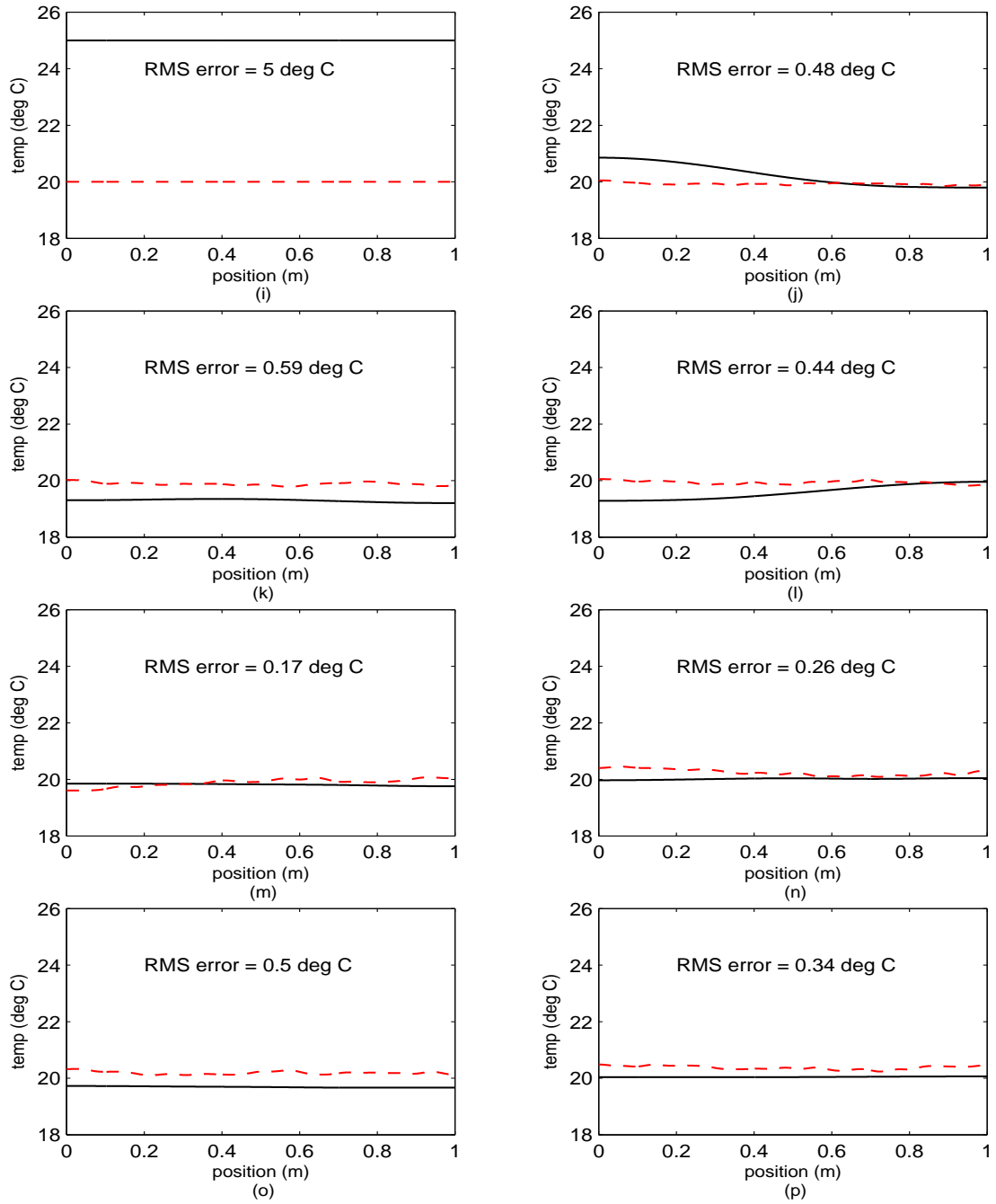


Figure 5.55 Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

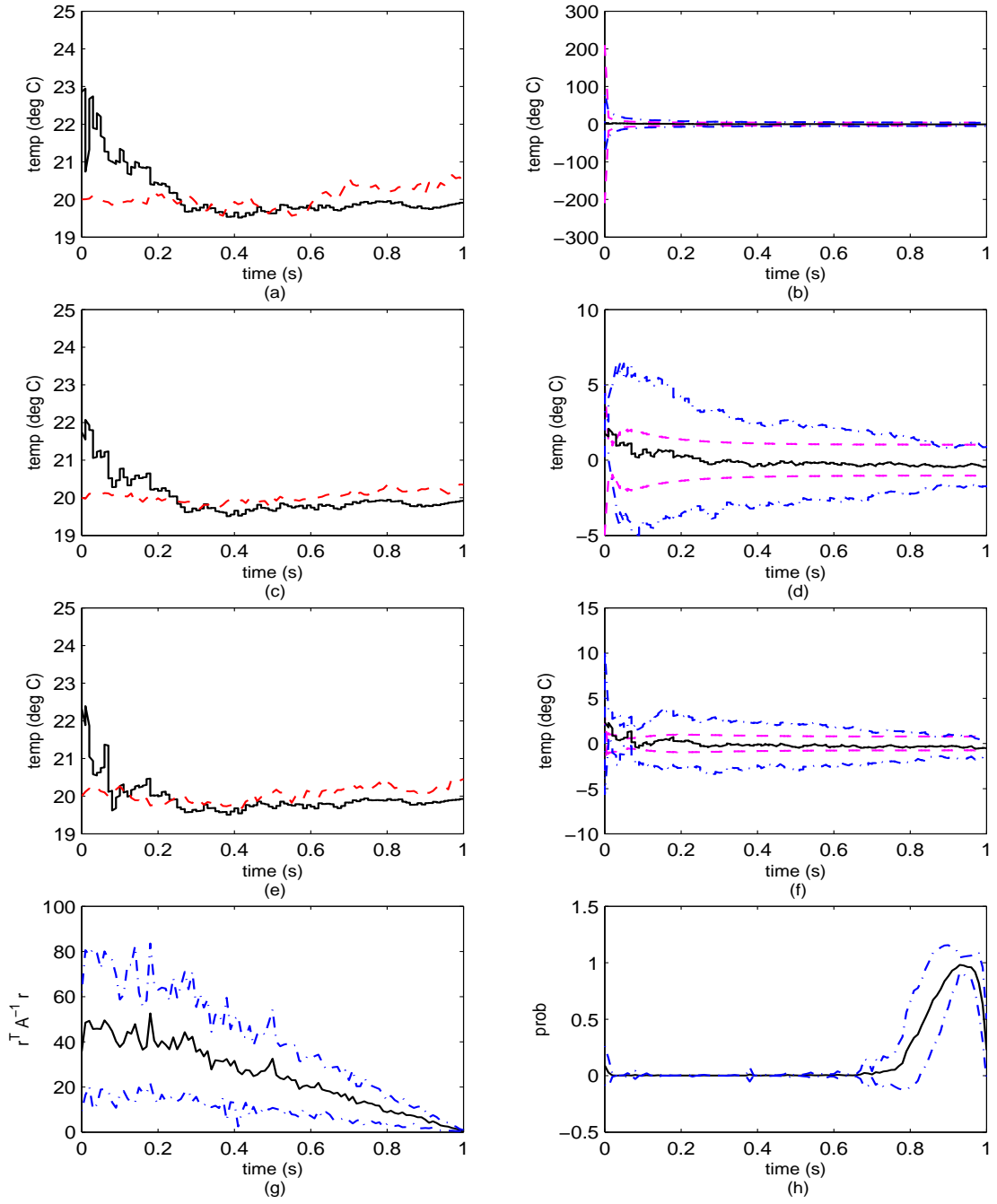


Figure 5.56 Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

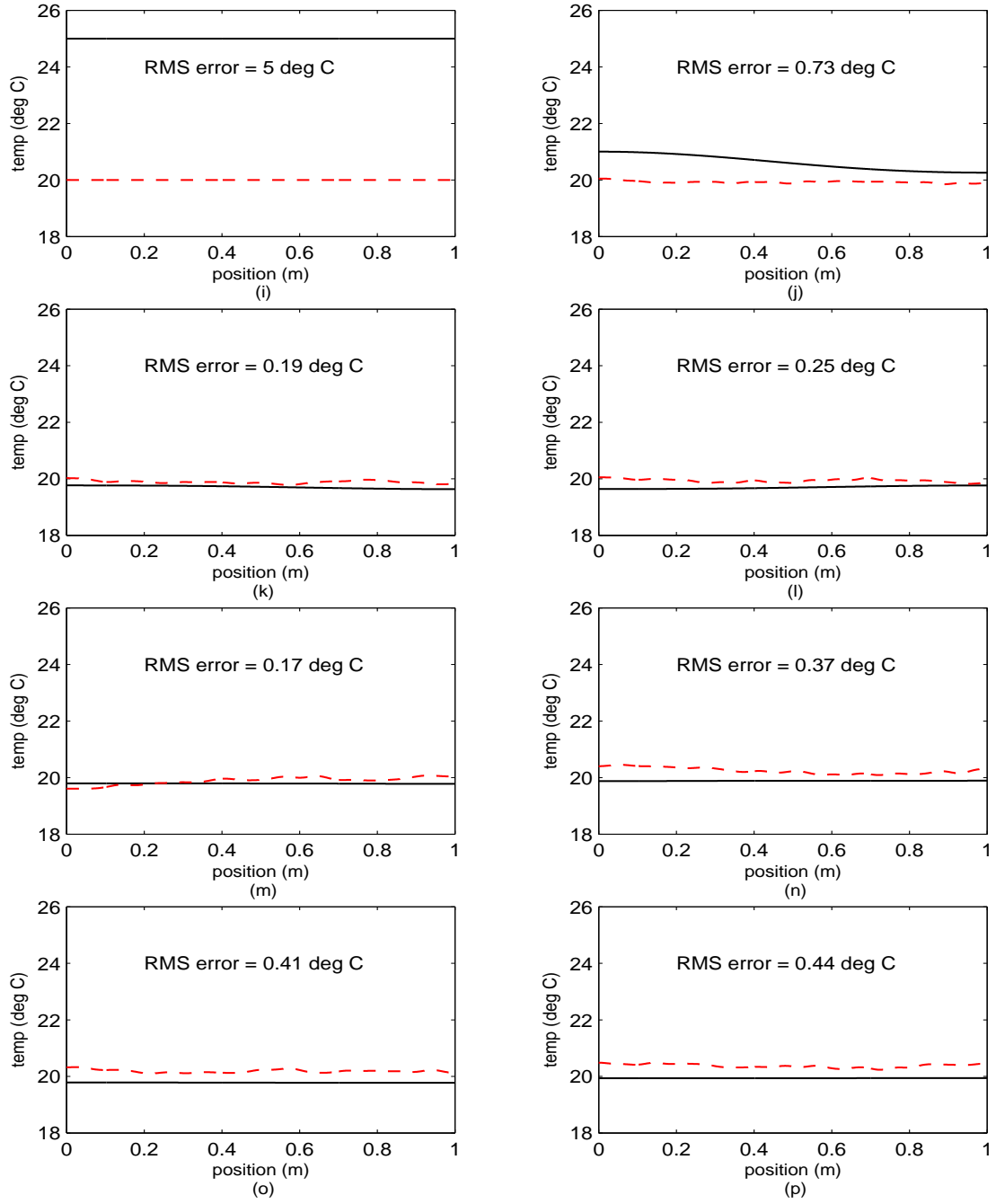


Figure 5.56 Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

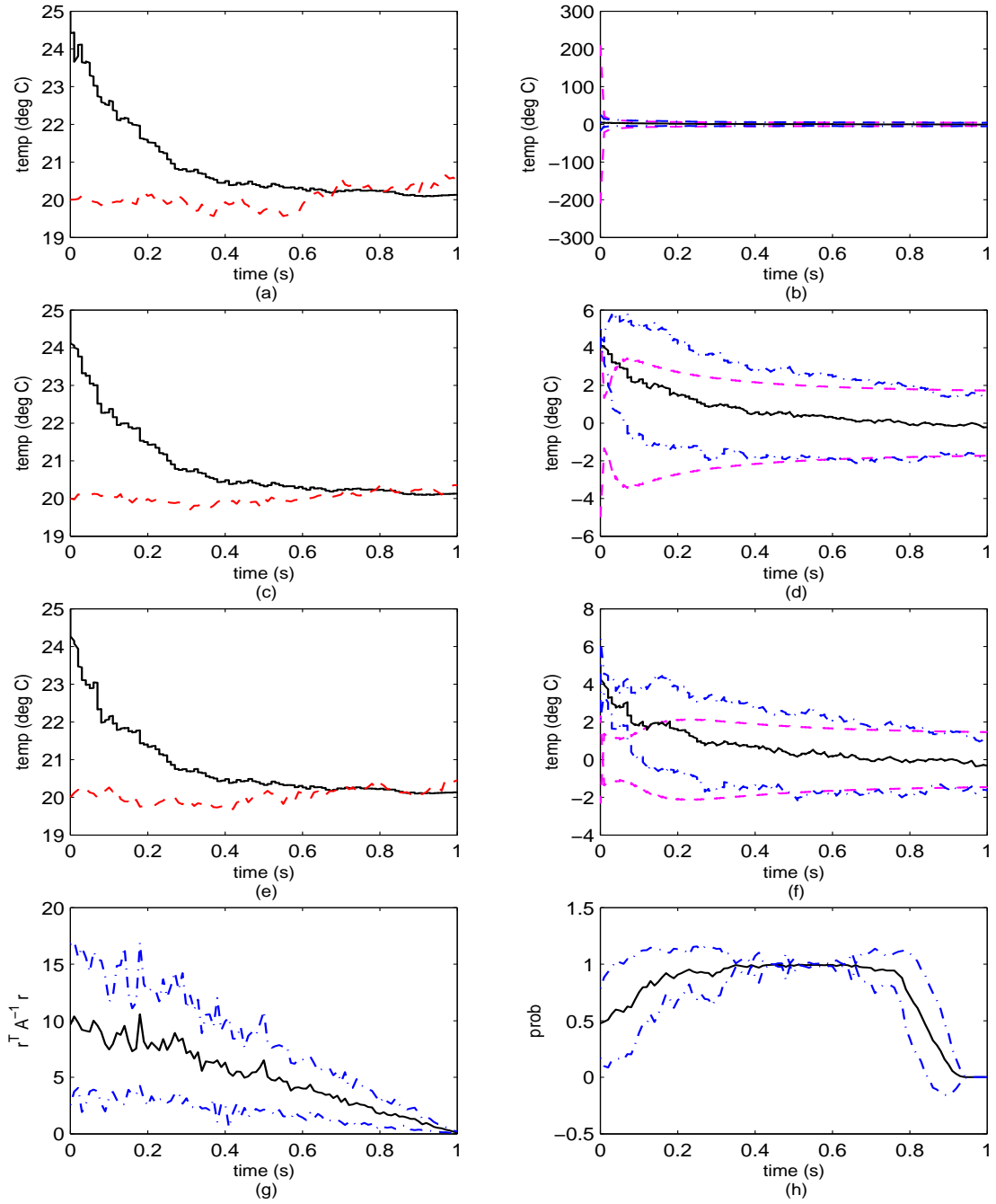


Figure 5.57 Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 4. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

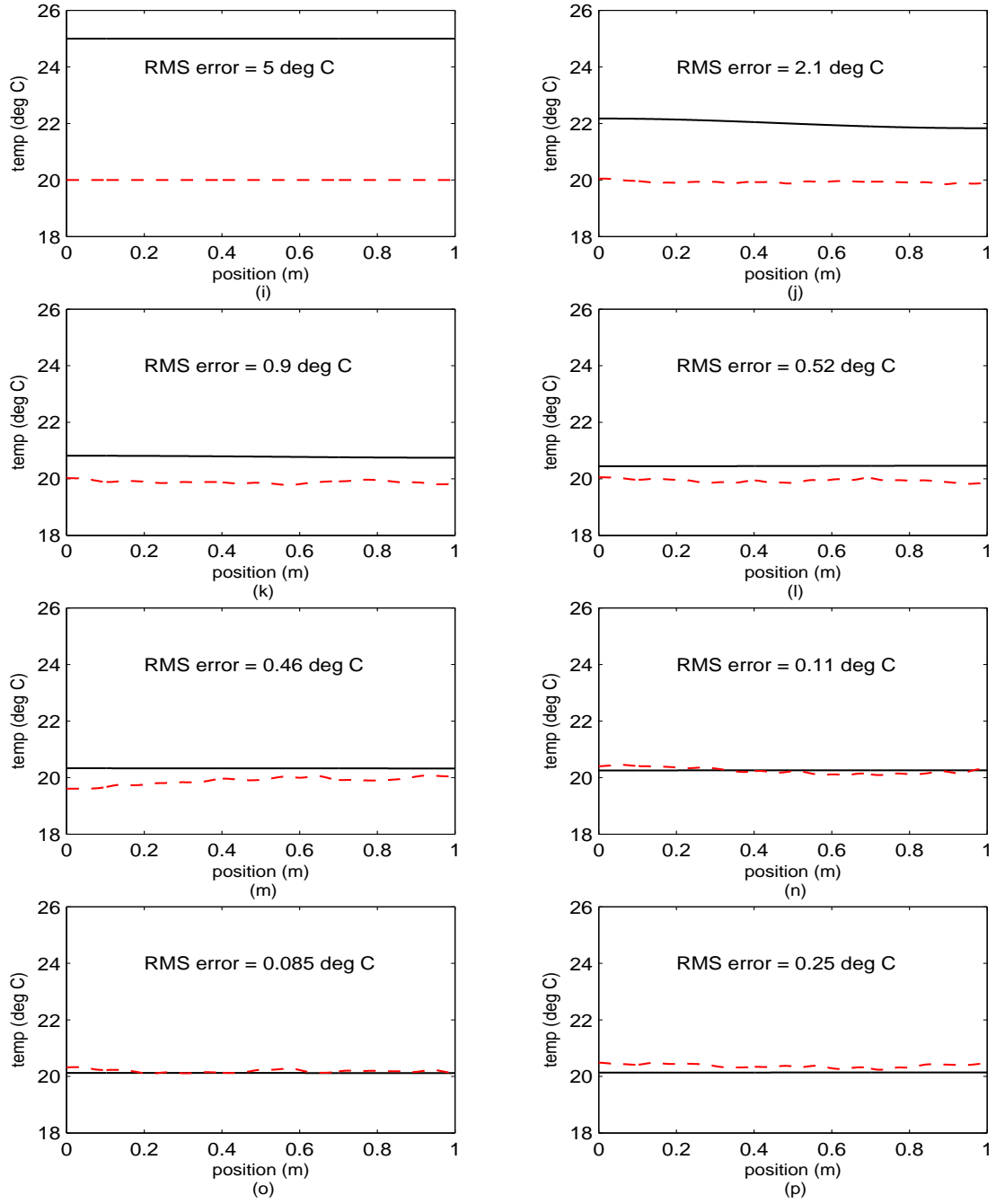


Figure 5.57 Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 4 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

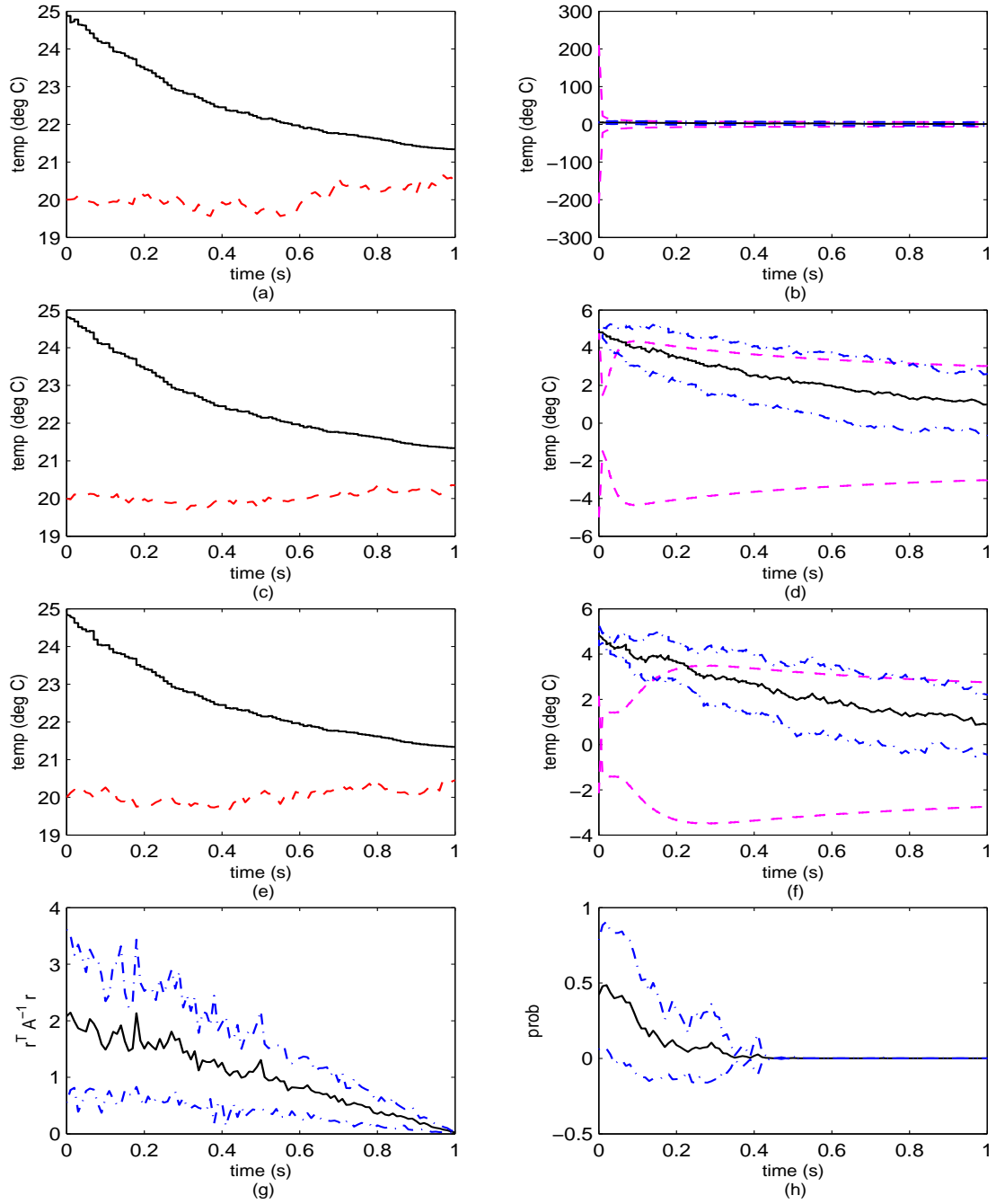


Figure 5.58 Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 5. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

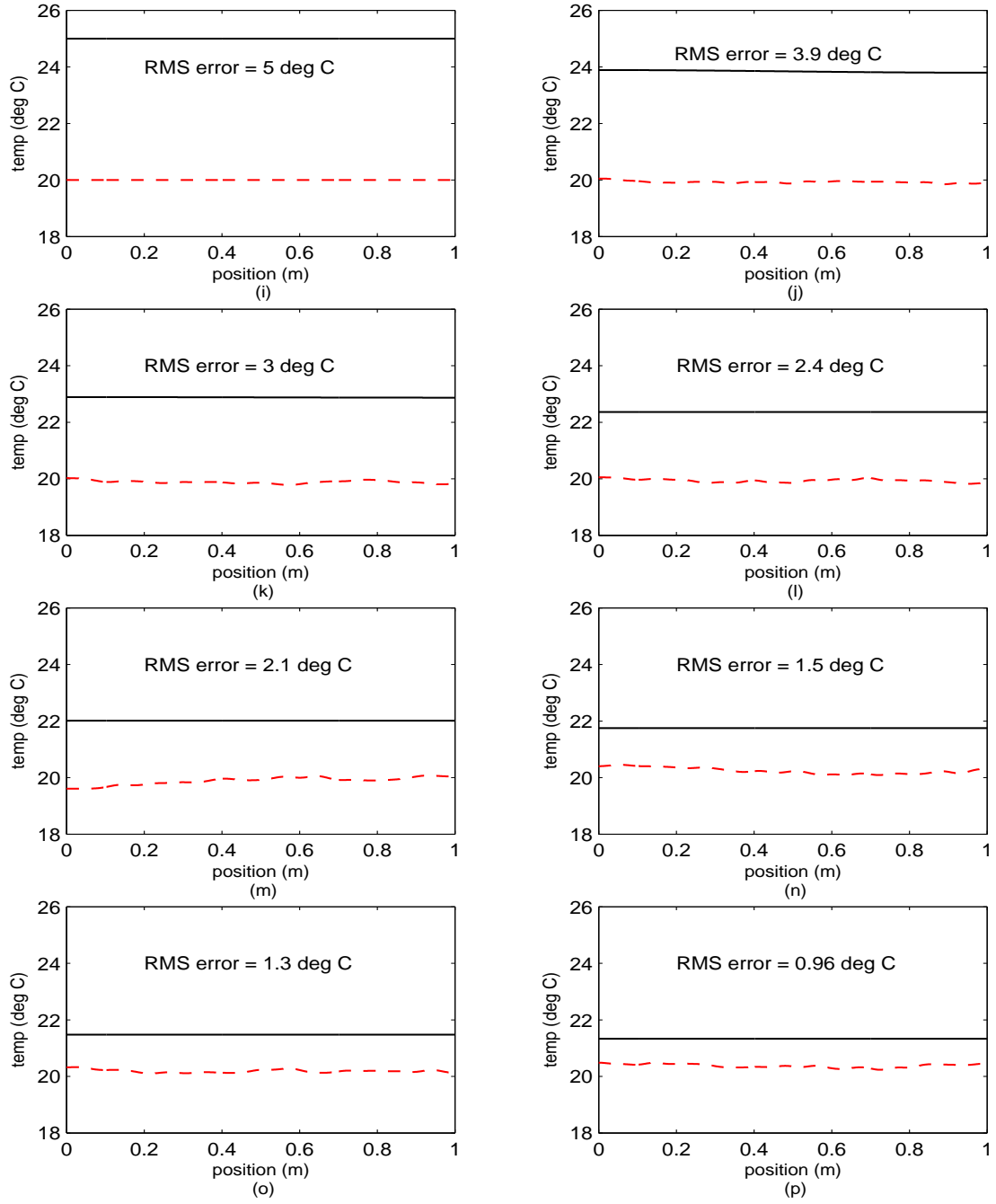


Figure 5.58 Simulation 4, Case 2 (decreasing R_{true}): Elemental Filter 5 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

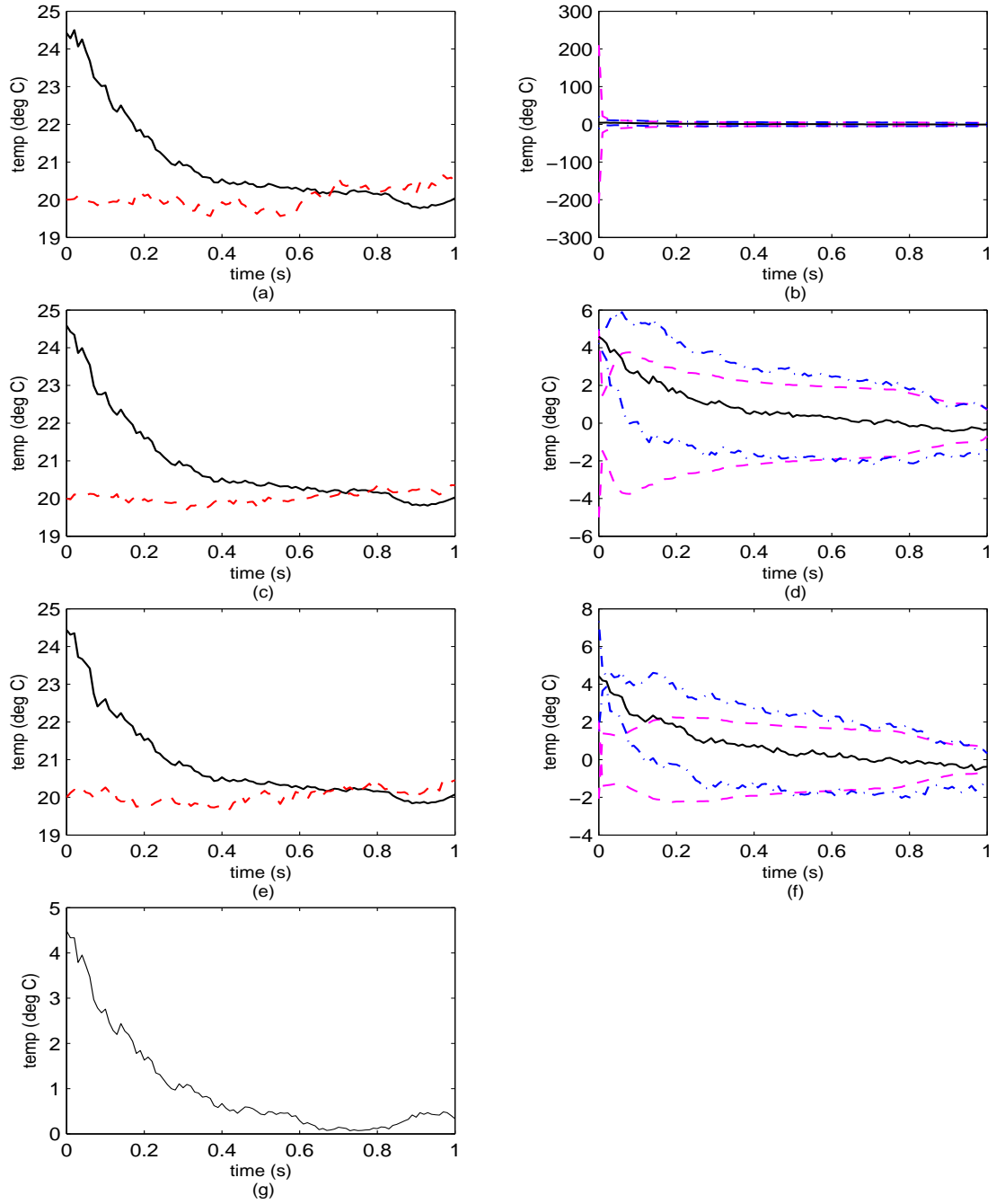


Figure 5.59 Simulation 4, Case 2 (decreasing R_{true}): Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

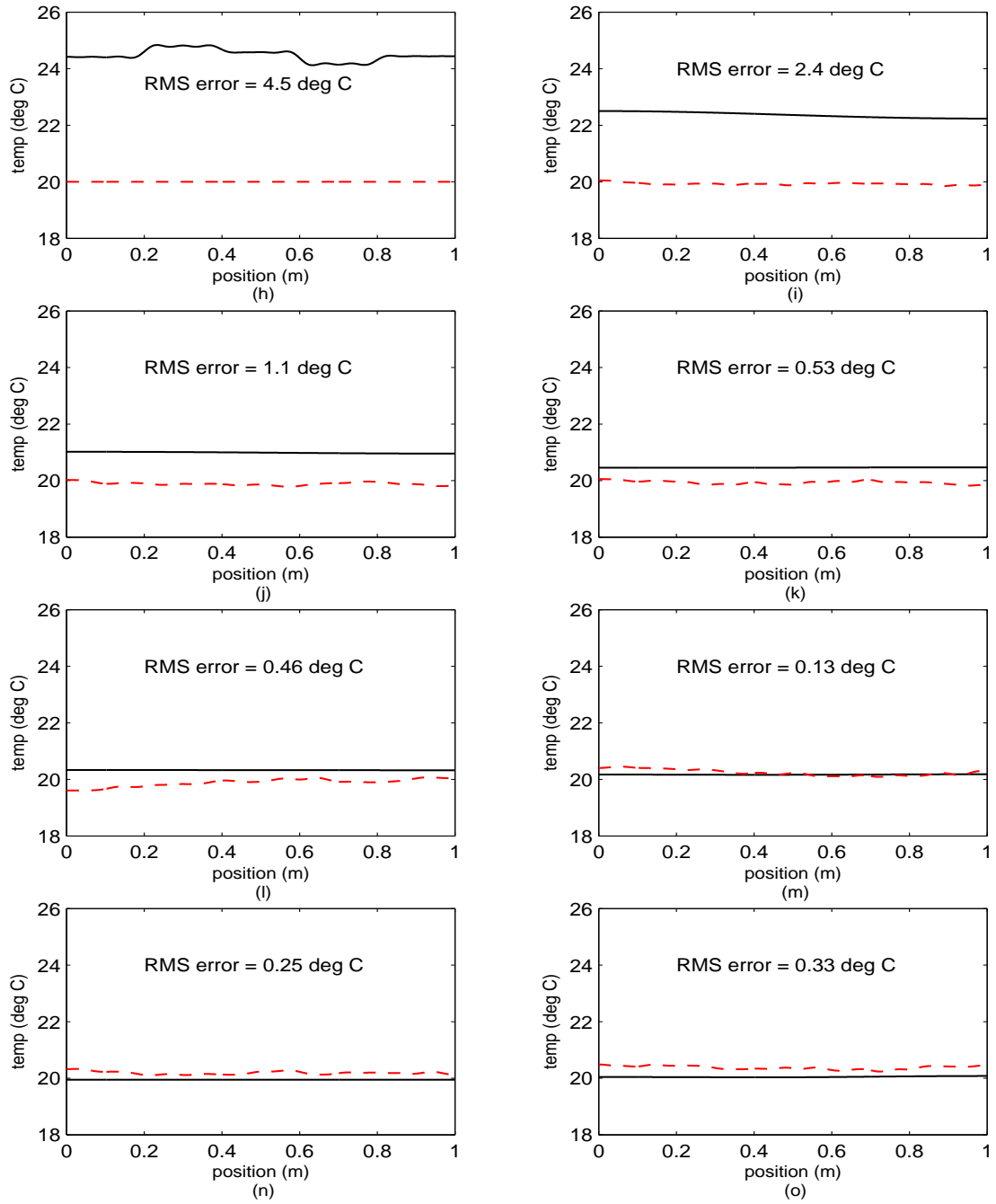


Figure 5.59 Simulation 4, Case 2 (decreasing R_{true}): Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

Filter	Q	R	\hat{x}_0	P_0
1		$0.04R_{\text{true}0} = 0.2$		
2		$0.2R_{\text{true}0} = 1$		
3	Q_{true}	$R_{\text{true}0} = 5$	25	25
4		$5R_{\text{true}0} = 25$		
5		$25R_{\text{true}0} = 125$		

Table 5.11 Simulation 5: Elemental Filter Parameters. $R_{\text{true}0}$ represents the initial value for R_{true} .

5.6 Simulation 5

In this simulation we investigate the MMAE's ability to respond to an abrupt change of the true measurement-corruption noise covariance, whereas in the last simulation the MMAE adjusted to a smooth linear change. We shall assume that we know the discrete set of the most likely R_{true} values that apply at any given time of interest. This known set is used to build a bank of five elemental filters paired to the five possible R_{true} values. The elemental filter design parameters are shown in Table 5.11. Once again we have used the truth value for Q . To be thorough, we conducted this simulation with a dynamics noise strength two, five, and ten times as large, with no significant change in the parameter estimation performance, just minor slowing of the response to the change, which is somewhat masked when Q is overestimated. Given the focus of this simulation, only the plots for the elemental filters for the truth level of dynamics noise strength are shown.

The true measurement-corruption noise covariance was abruptly changed during the simulation as follows: in the first third of the scenario, filter 3 presents the best model as can be seen in Figure 5.60. In the second third, filter 4 is the best hypothesized value, and finally, filter 2 matches the best during the final third. As expected, only one elemental filter achieved the ideal likelihood quotient, $L_k(t_i) = \mathbf{r}_k^T(t_i) \mathbf{A}_k^{-1}(t_i) \mathbf{r}_k(t_i)$, of 5 during each period. For the first, second, and third periods, the average likelihood quotient was 5 for elemental filters 3, 4, and 2,

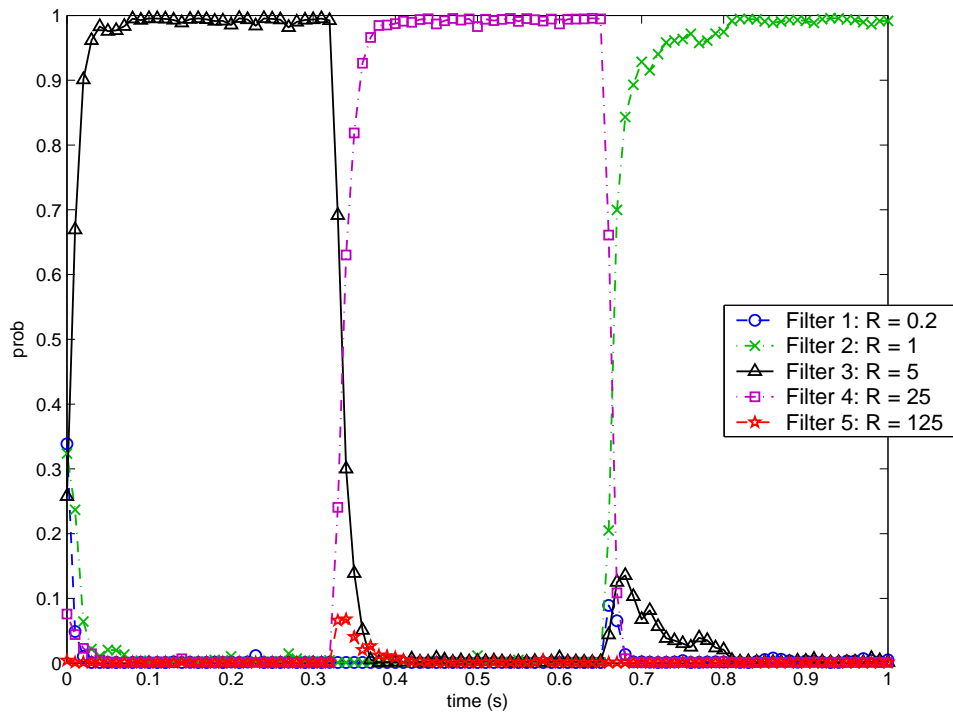


Figure 5.60 Simulation 5: Hypothesis conditional probability.

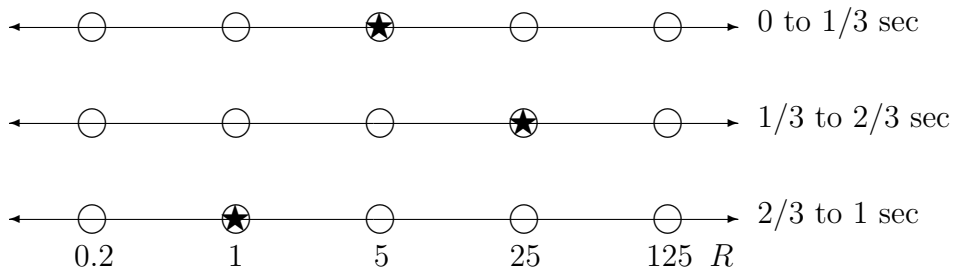


Figure 5.61 Simulation 5: Abruptly changing measurement-corruption noise covariance. Legend: \bigcirc elemental filter; \star true parameter within the filter bank. (The filter spacing is nonlinear for illustration purposes.)

as shown in Figures 5.64(g), 5.65(g), and 5.63(g), respectively. This progression is further illustrated in Figure 5.61.

The clean, fast convergence displayed during the first change from elemental filter 3 to filter 4 can be explained by noting that, as the true R experienced a positive step increase, the assumptions upon which filter 3 was constructed were now being

Period:	First	Second	Third
Truth:	R_{true_0}	$5R_{\text{true}_0}$	$R_{\text{true}_0}/5$
Filter 1 ($R_1 = R_{\text{true}_0}/25$)	125	625	25
Filter 2 ($R_2 = R_{\text{true}_0}/5$)	25	125	5
Filter 3 ($R_3 = R_{\text{true}_0}$)	5	25	1
Filter 4 ($R_4 = 5R_{\text{true}_0}$)	1	5	1/5
Filter 5 ($R_5 = 25R_{\text{true}_0}$)	1/5	1	1/25

Table 5.12 Simulation 5: Expected likelihood quotients. R_{true_0} is the initial value for R_{true} . The correctly modeled elemental filter will have a likelihood quotient of about five.

violated and thus the probability flow to filter 4 was rapid. The second probability flow transition, which was prompted by a large decrease in the true noise covariance, was less crisp than the first transition. While the probability for filter 4 dropped off rapidly, filter 3 initially earned a small share of the probability during this second transition since it was not immediately clear that the measurement quality change was real versus spurious. These observations match up well with the analysis based on the likelihood quotients given in Simulation 4; there we noted that the MMAE responds more quickly and adroitly to a positive change in R_{true} than it does for a negative change in R_{true} . Thus, we would expect the same general performance to a step change in R .

Note that the *shape* of the likelihood quotient plot is the same for all of the elemental filters — only the magnitude has changed in plots (g) of Figures 5.62 through 5.66. The expected magnitudes for each of the elemental filters in Table 5.12 can be found by using the steady state likelihood quotient for the k th elemental filter at time t_i derived in Section 5.1 and repeated here for our convenience

$$E\{\mathbf{L}_k(t_i)|_{t_i=t_{ss}}\} \cong \frac{R_{\text{true}}}{R_k} M \quad (5.10)$$

When the hypothesis is correct, $R_k = R_{\text{true}}$, and then the right-hand side of Equation (5.10) becomes M . The interesting cases occur when $R_j \neq R_{\text{true}}$. For example, when R_{true} is “under”-modeled by a factor of five, then the likelihood quotient is five times bigger than for a properly modeled filter. The likelihood quotient of 25 occurs during the three times of the simulation when the truth R is five times the assumed value in the elemental filter. Just the opposite occurs when the truth R is one fifth of the assumed value in the filter. This also happens three times during the simulation as noted in the table.

While the likelihood quotient is a good barometer for how well an elemental filter matches the real environment, a quick review on how the hypothesis conditional probability is calculated, see Equations (2.43) through (2.46) on pages 2-29 and 2-30, reveals that there is another term that we should consider. The leading coefficient of the Gaussian density, the β -term shown in Equation (2.45), plays an important role when the residuals for multiple filters are approximately equal. Consider the scalar measurement case: for approximately equal residuals, the elemental filter with the smaller filter-computed residual covariance is “preferred” by the MMAE.

In the end, we anticipate that the innate capability of the MMAE to adapt to an unknown noise environment will result in an improved state estimate. As we inspect the blended filter plots in Figure 5.67 on page 5-153, we note that the majority of the plots appear to have three distinct regions — this is due to the abruptly changing truth R . In each of the time periods in which a different R_{true} was in force, the MMAE quickly identified the best elemental filter; this leads to a quality state estimate for that elemental filter for that time period and consequently, it predominantly determined the blended filter results given that it gathered approximately all of the probability during that particular time slot. The initial R_{true} gave reasonably good measurements and thus the MMAE was able to handle the initial state estimate bias with ease. When the truth measurement-corruption noise covariance increased by a factor of five in the second period, the MMAE quickly “decided” that the

measurement noise covariance had increased somewhat and filter 4 soon received the bulk of the probability. At the end of the first period (from 0 to 0.33 sec), the RMS error in plot (g) of Figure 5.67 was quite small, owing to the fact that R_{true} was small. In the second period (from 0.33 to 0.67 sec), R_{true} was five times as large, hence it should not come as a surprise that the mean-squared error would increase by a factor of five initially. The improved measurement quality during the third period brought the RMS error back down to its lowest levels. The two transitions can be seen clearly in plots (g) and (h) of Figures 5.63, 5.64, and 5.65 for elemental filters 2, 3, and 4 respectively.

For the two elemental filters designed to model poor quality measurements, elemental filters 4 and 5, poor initial transients create the unusual circumstance seen in Figures 5.65 and 5.66, plots (d) and (f). As noted earlier, the adequacy of the initial state covariance can be checked by inspecting plots (b), (d), and (f); the initial error should be within the 1σ (i.e., one standard deviation) bounds created by the initial state covariance. If this is not true, then convergence is greatly hampered since the filter has been told that its initial condition errors are much smaller than they really are.

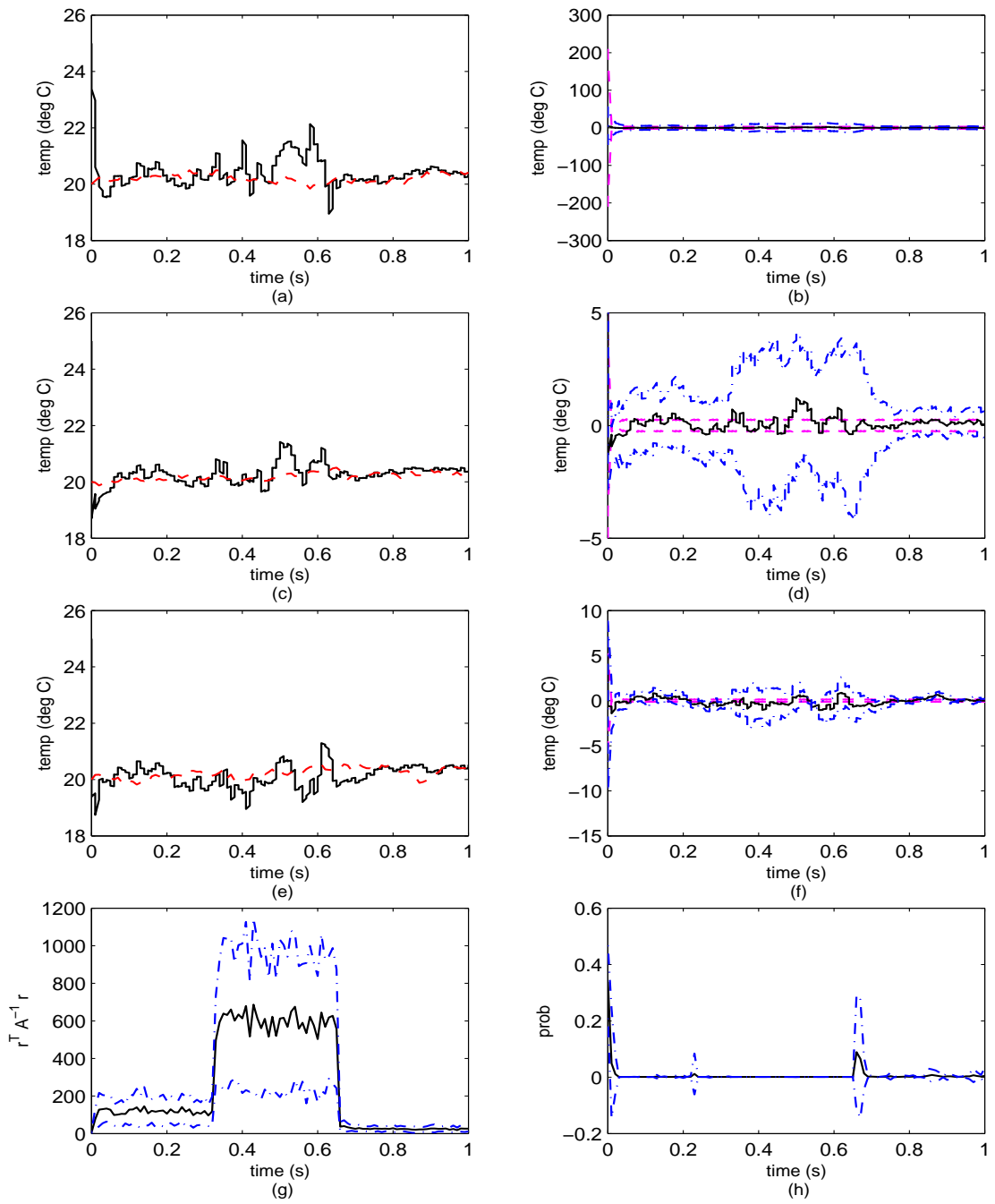


Figure 5.62 Simulation 5: Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

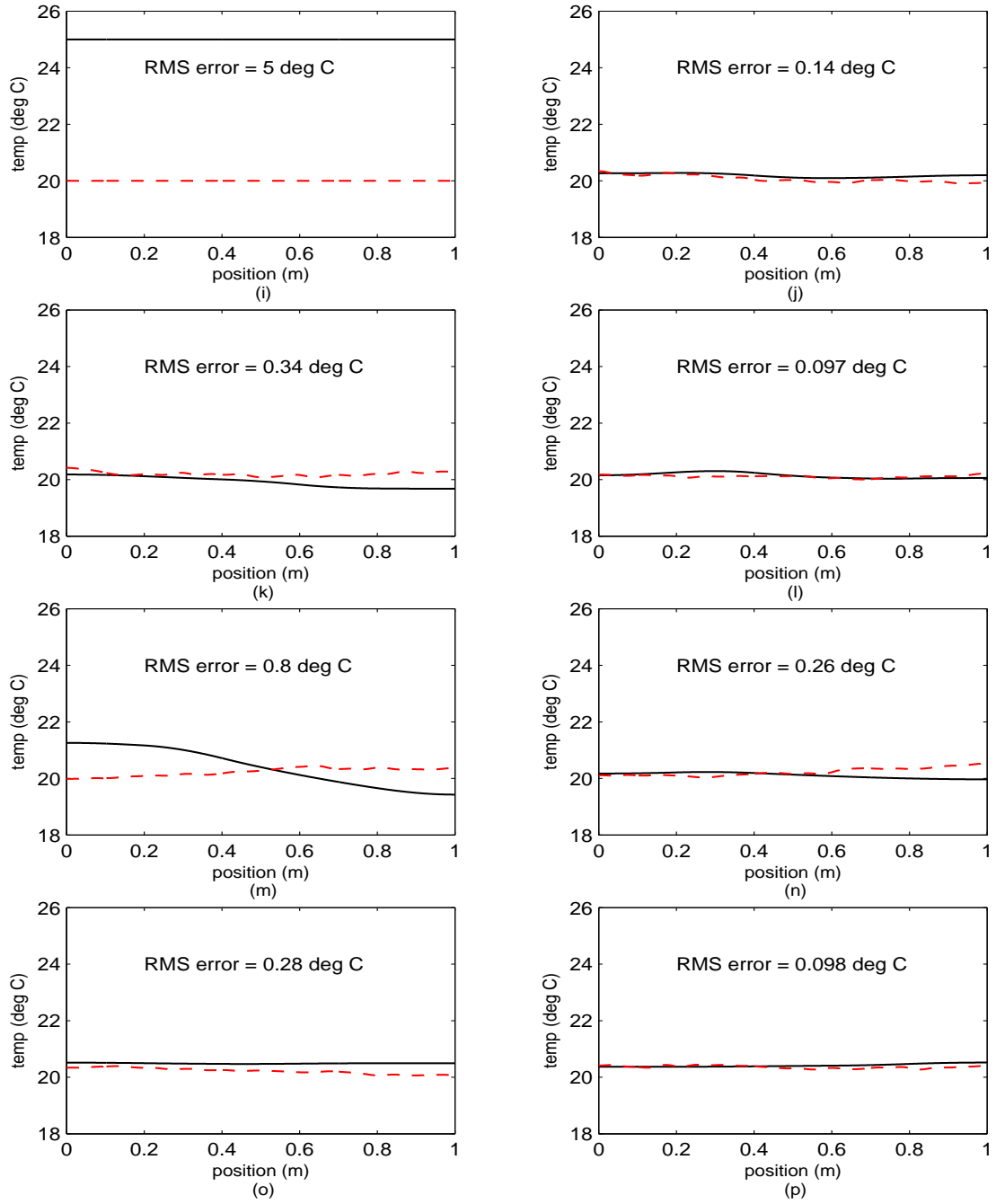


Figure 5.62 Simulation 5: Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

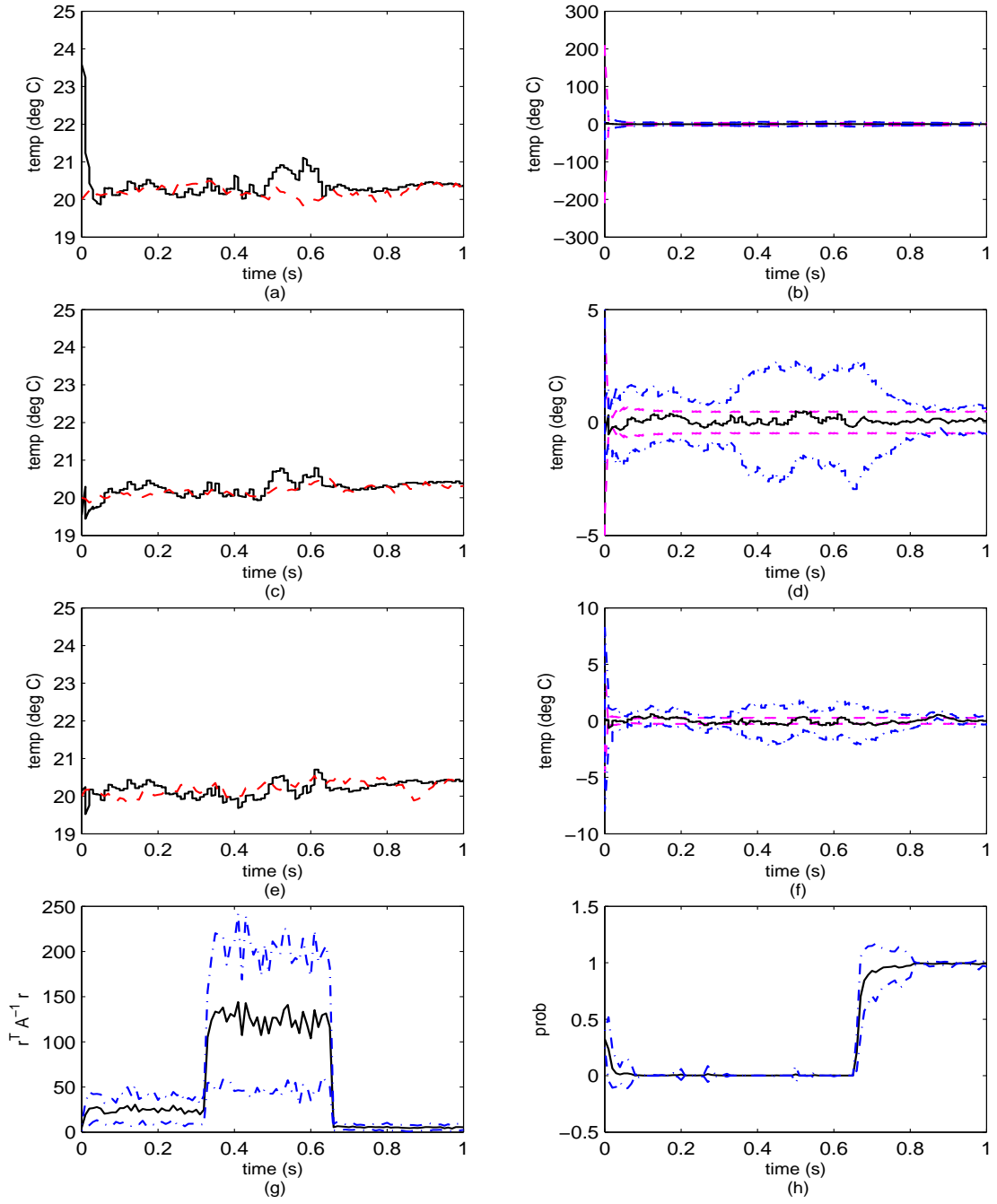


Figure 5.63 Simulation 5: Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

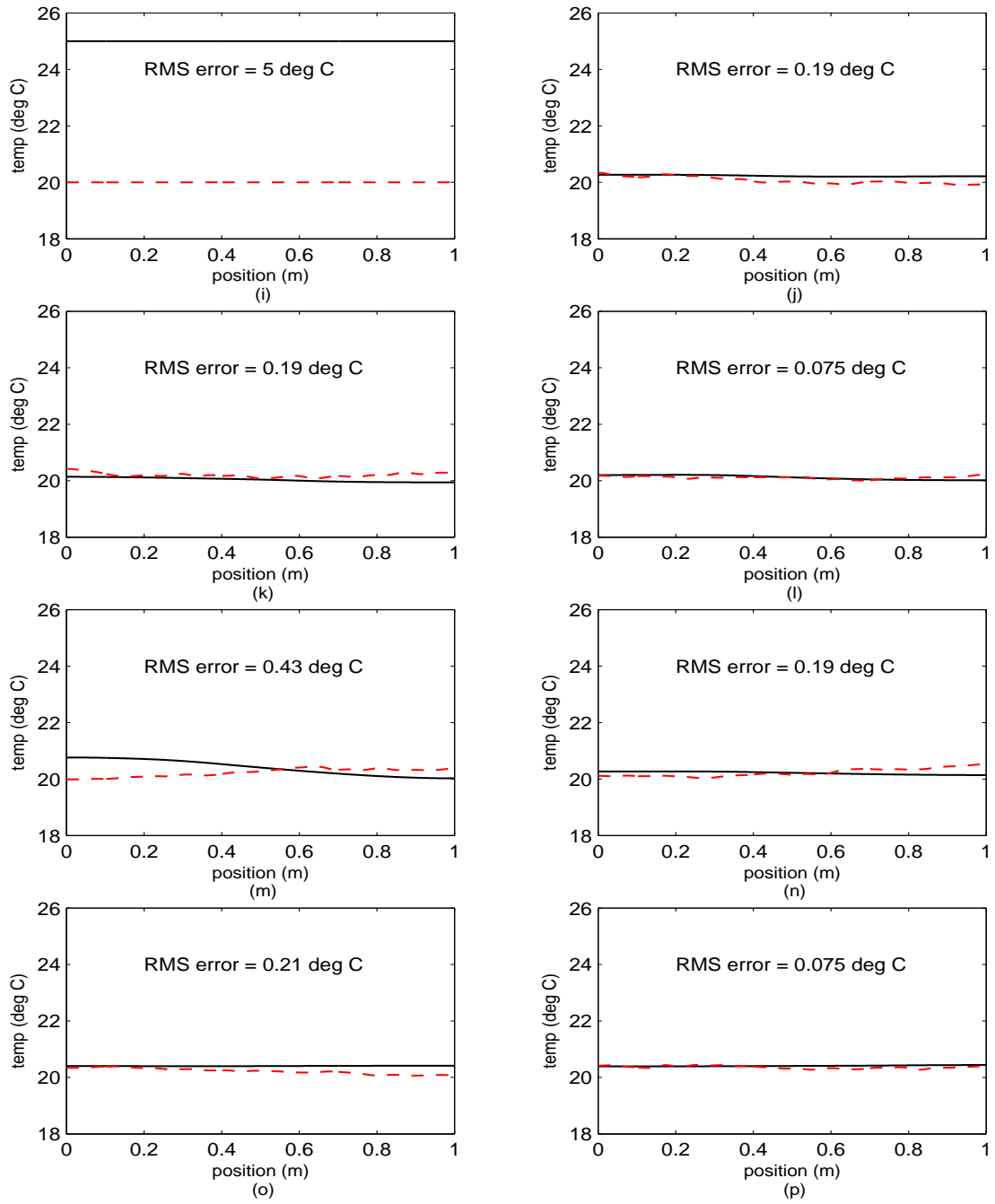


Figure 5.63 Simulation 5: Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

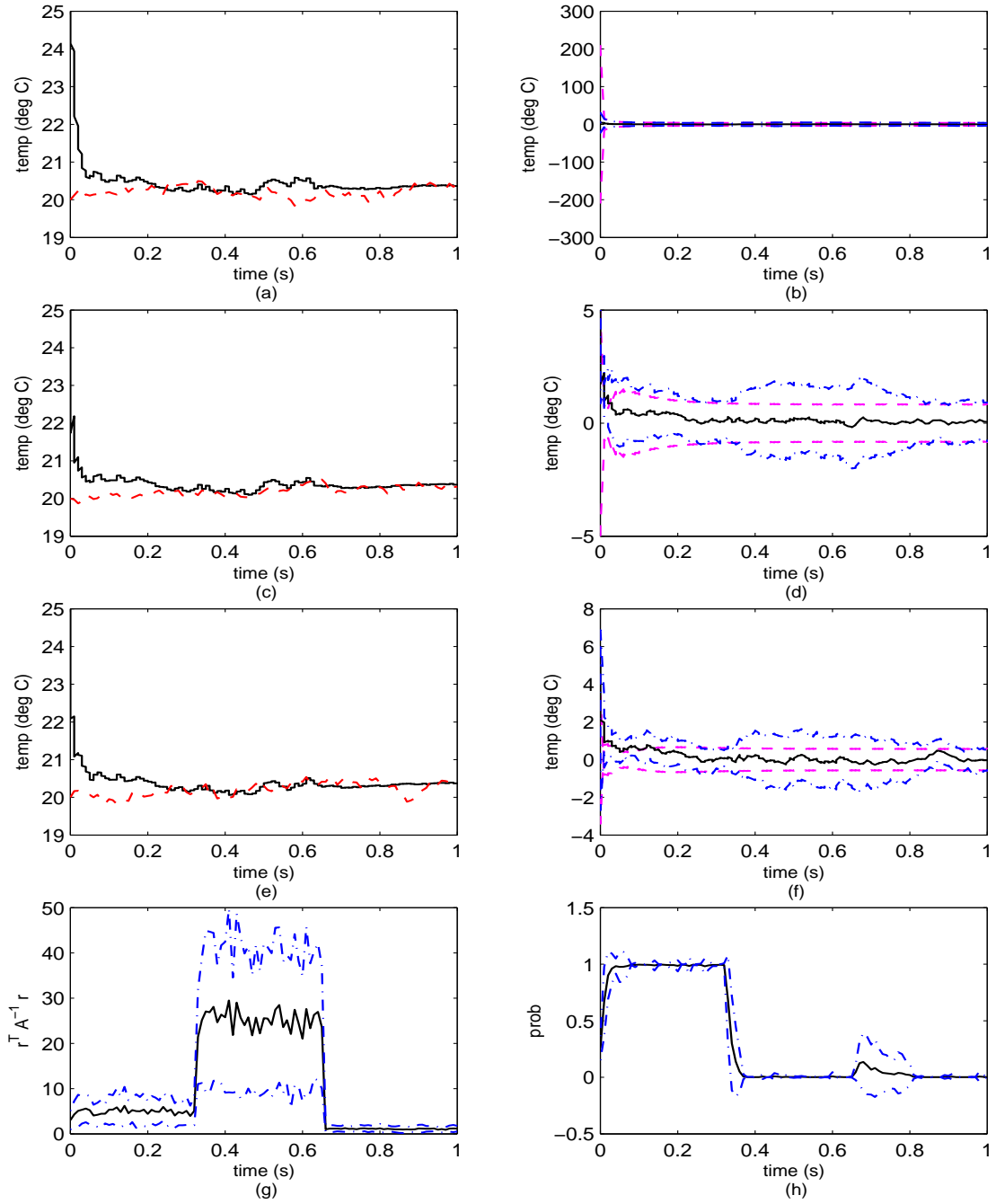


Figure 5.64 Simulation 5: Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

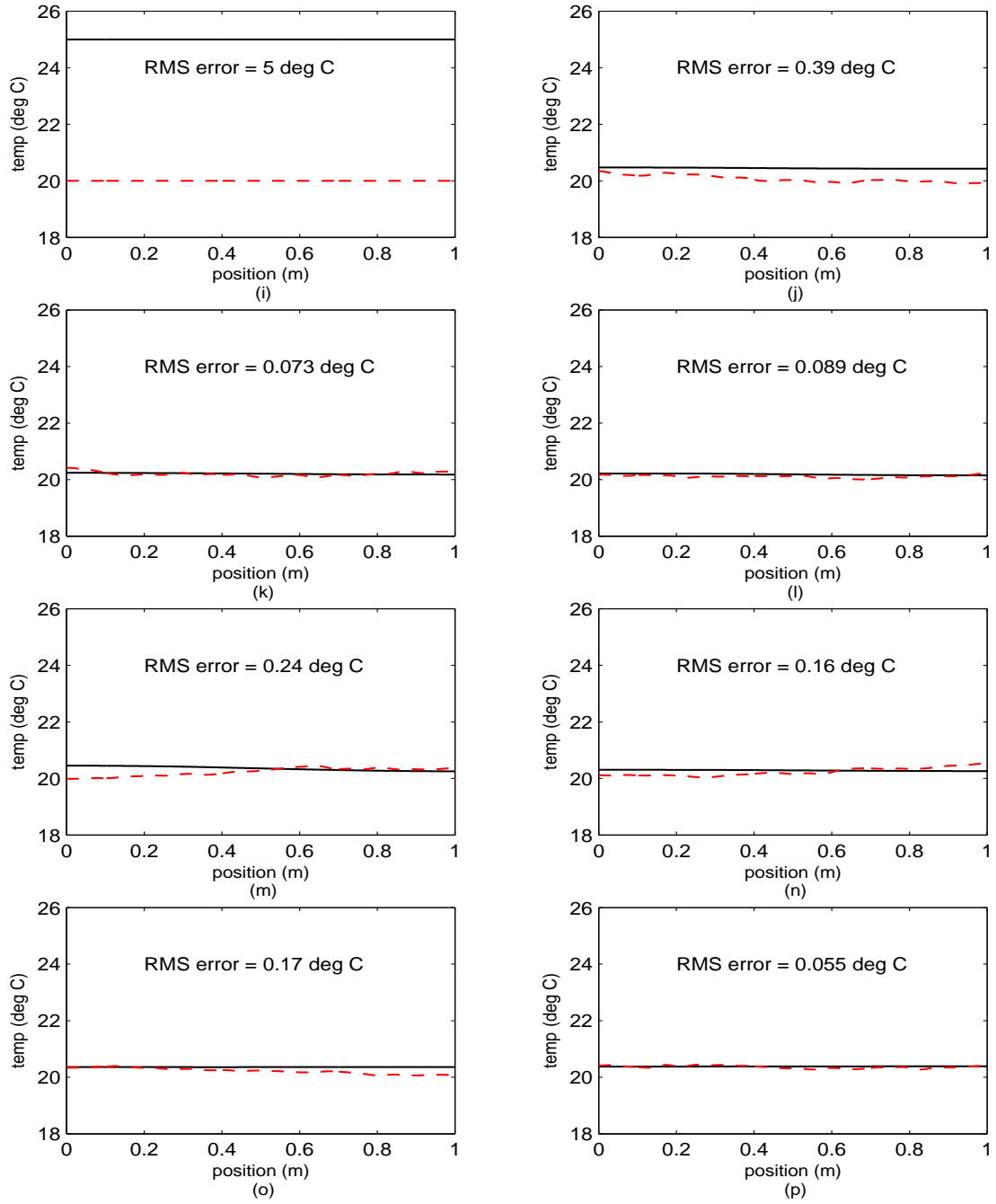


Figure 5.64 Simulation 5: Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

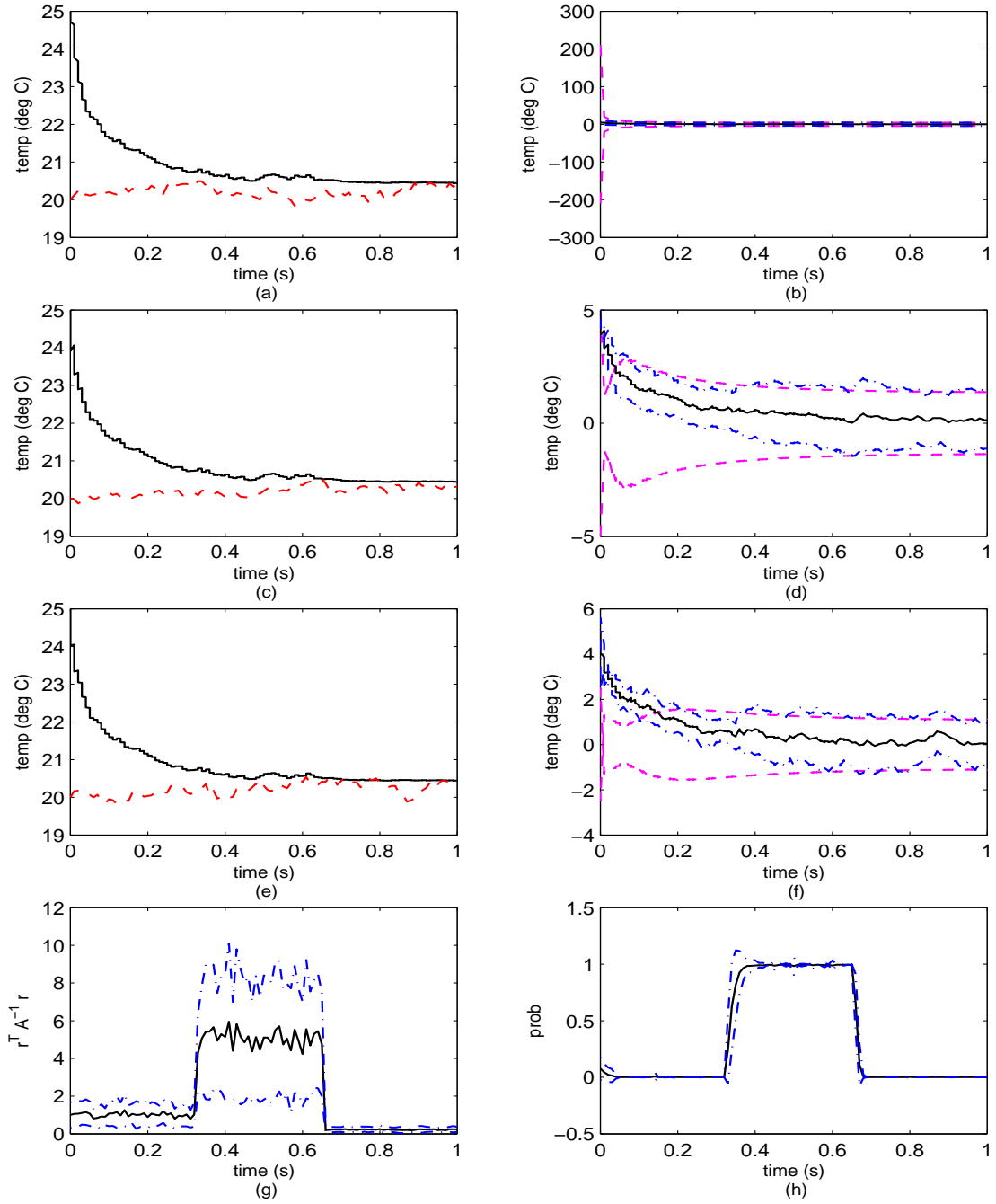


Figure 5.65 Simulation 5: Elemental Filter 4. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

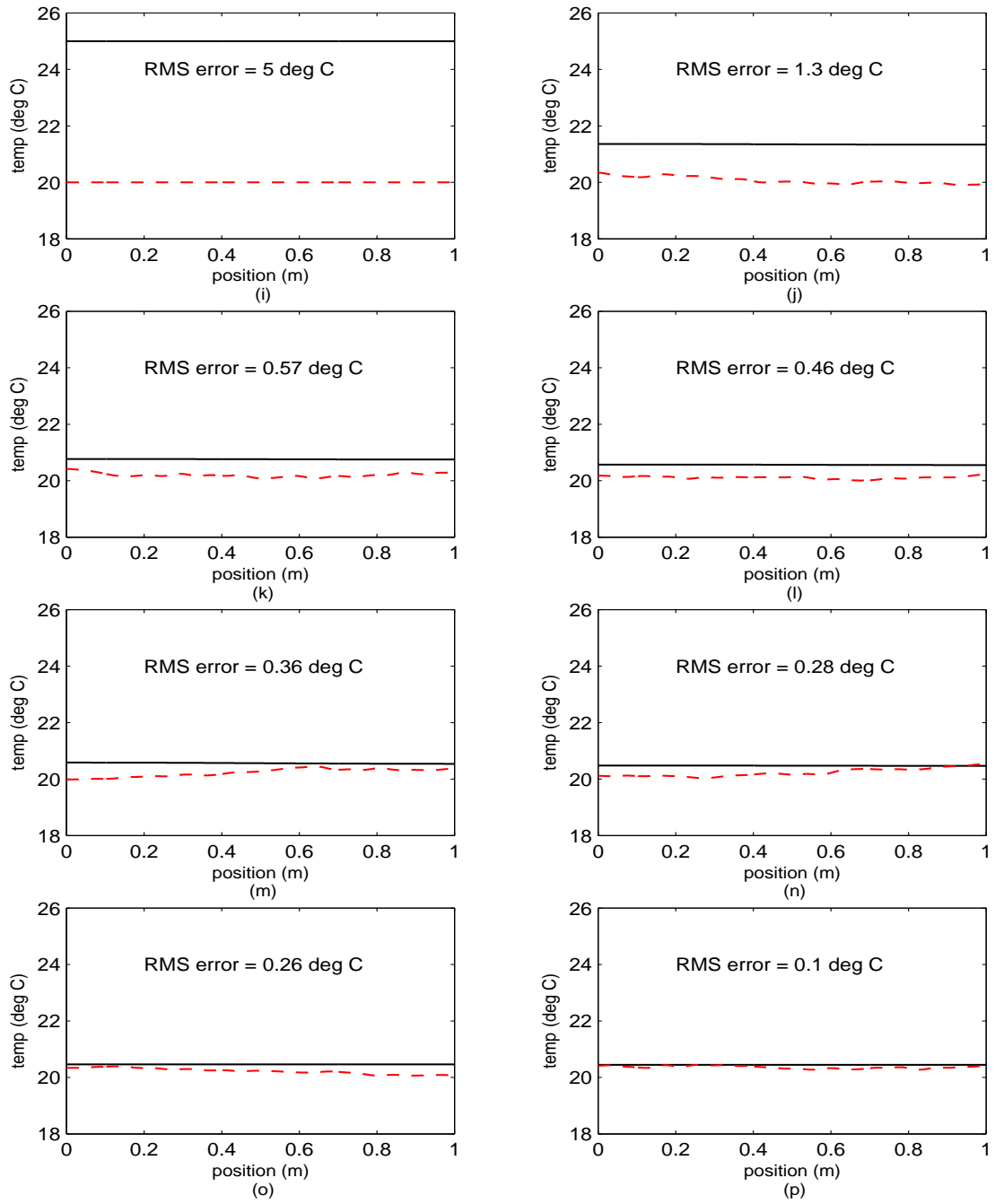


Figure 5.65 Simulation 5: Elemental Filter 4 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

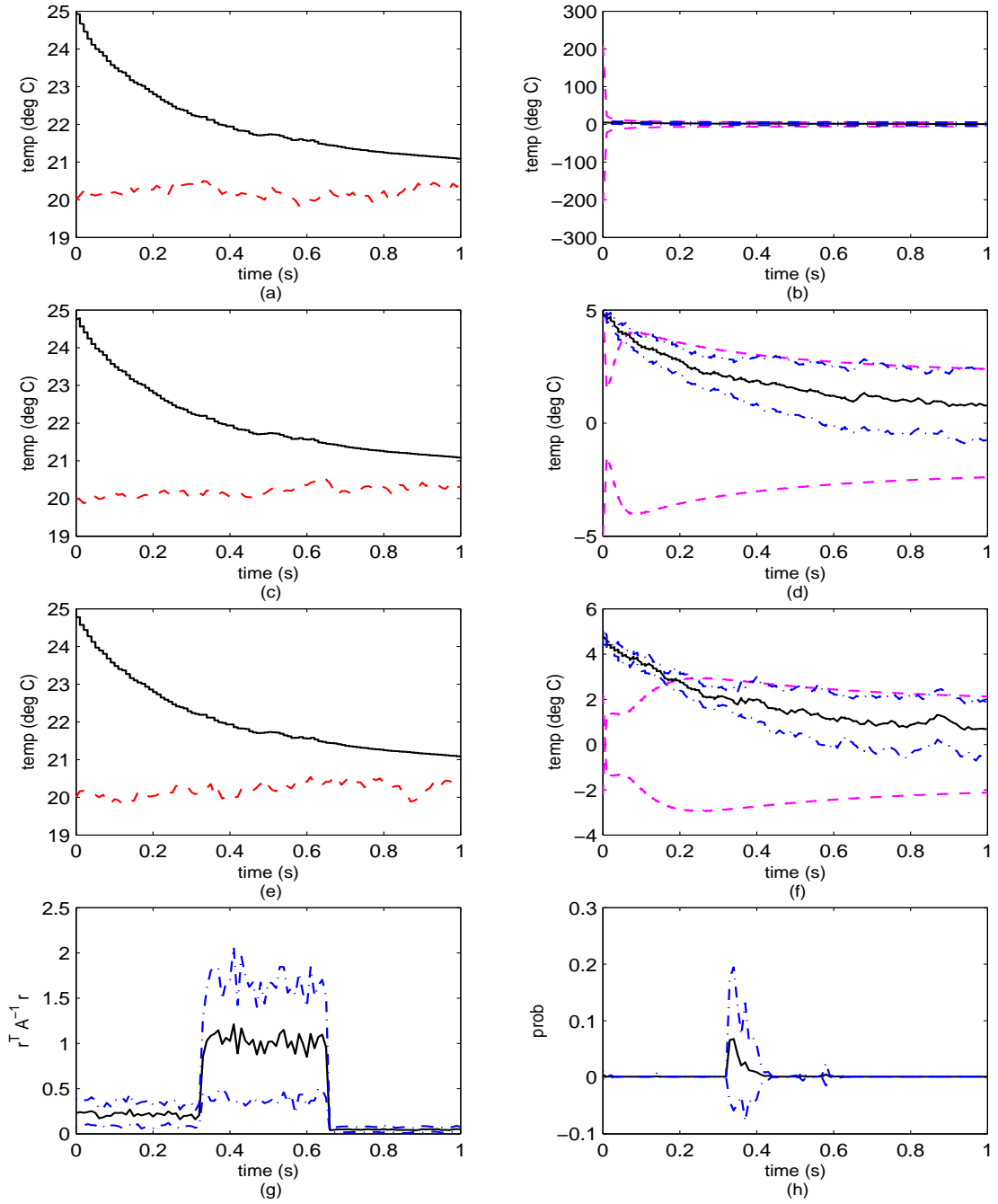


Figure 5.66 Simulation 5: Elemental Filter 5. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

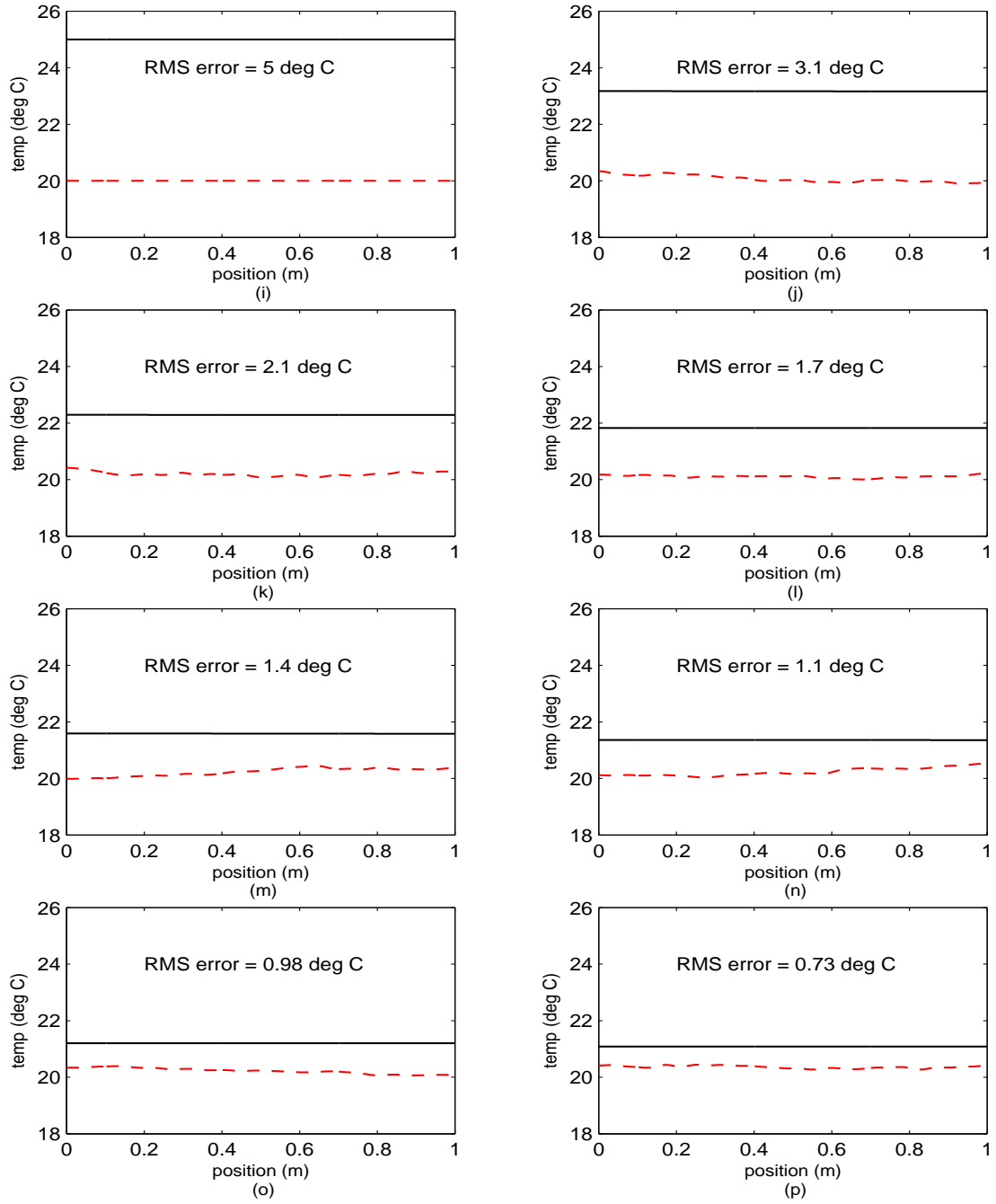


Figure 5.66 Simulation 5: Elemental Filter 5 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

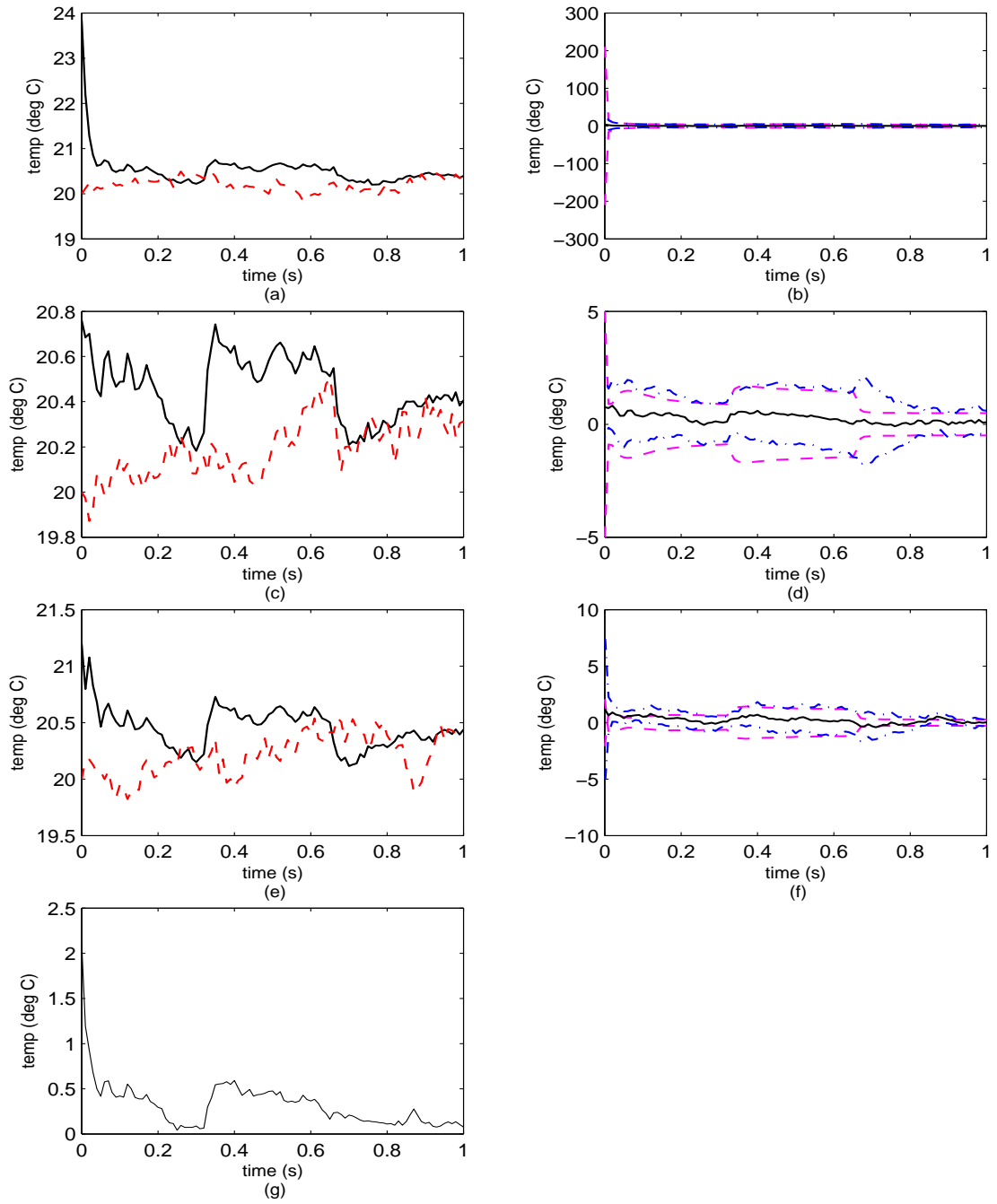


Figure 5.67 Simulation 5: Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

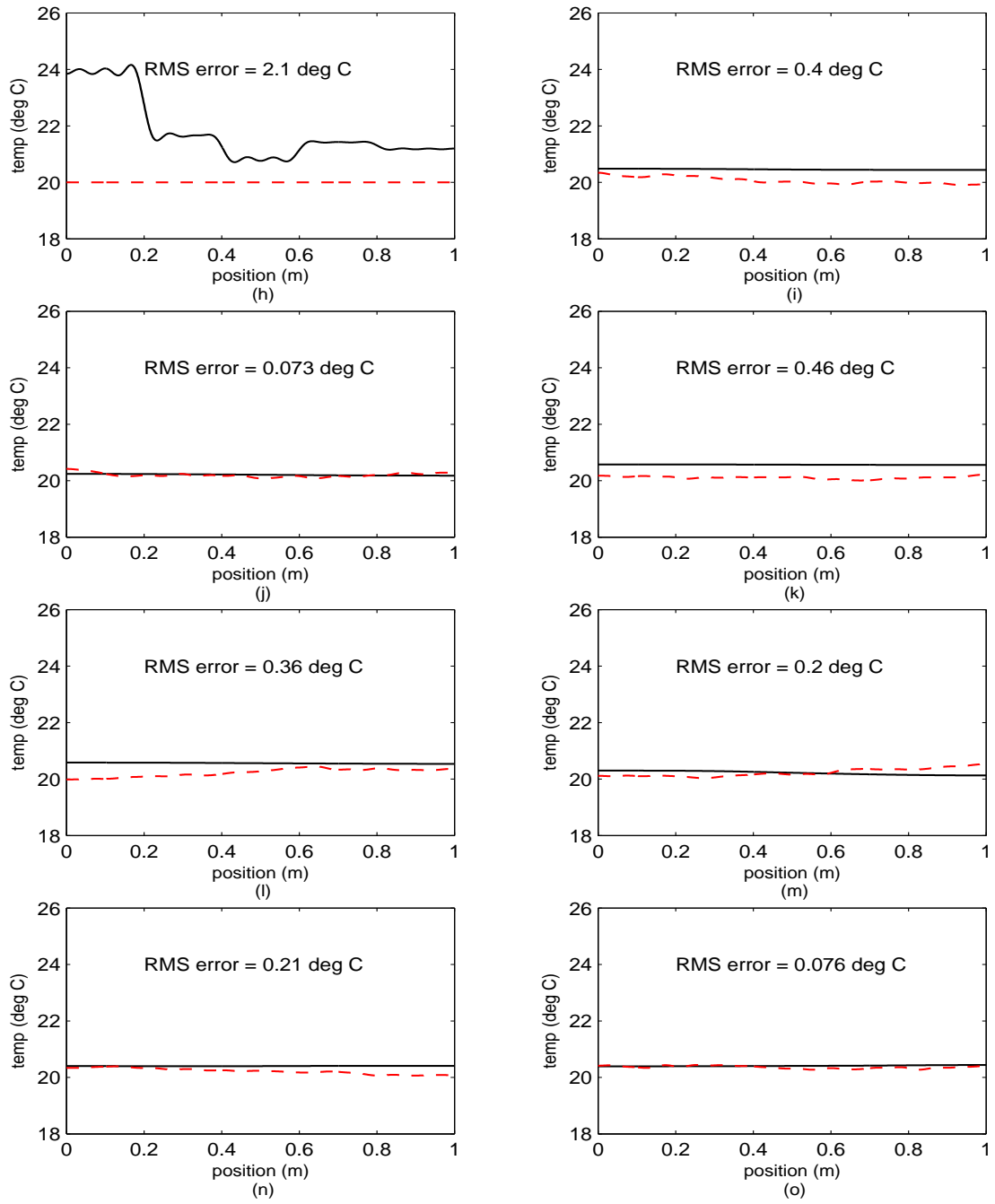


Figure 5.67 Simulation 5: Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

Q	R	\hat{x}_0	κ
5	5	20	0.86 (aluminum)

Table 5.13 Simulation 6: Truth Parameters

Filter	Q	R	\hat{x}_0	P_0	κ
1					0.011 (granite)
2					0.12 (cast iron)
3	Q_{true}	R_{true}	25	25	0.86 (aluminum)
4					1.14 (copper)
5					1.71 (silver)

Table 5.14 Simulation 6: Elemental Filter Parameters. The diffusivity constants are taken from Boyce and DiPrima [26].

5.7 Simulation 6

In this final simulation we examine the system identification capabilities of the MMAE. Some of the important truth system parameters are given in Table 5.13; take note that the thermal diffusivity constant is no longer 1, but has been programmed to represent an aluminum rod. We created five elemental filters assuming that we know the true noise environment statistics, i.e., zero-mean Gaussian stochastic process with known covariance matrices and a 5 °C bias in our initial state (temperature) estimate. Each of the five elemental filters is programmed with a distinct choice for the thermal diffusivity constant, κ , corresponding to a known thermal diffusivity for a specific material as shown in Table 5.14. Physically, as diffusivity increases, resistance to heat flow decreases. A lower diffusivity means that heat flows (or diffuses) more slowly through the material. In other words, materials with “high” diffusivity constants achieve thermal equilibrium faster than materials such as granite. Notice how closely spaced the diffusivity constants (the parameters in this case) are for aluminum, copper, and silver.

By introducing a purposeful excitation signal into the system (in this case, heat applied to the left end of the rod), we increased the identifiability of the unknown

parameter dramatically and we were able to perform system identification at a very fine level. Since we were introducing this excitation to the system, we also gave the elemental filters a very good (but not exact) approximation of the heating profile. The true heating profile at the left end of the rod is⁹

$$u_{\text{true}}(t_i) = \begin{cases} \frac{0.01+t_i}{0.1} \eta v, & 0 \leq t_i \leq 0.09 \\ \eta v, & 0.1 \leq t_i \leq 0.75 \\ 0, & 0.76 \leq t_i \leq 1 \end{cases} \quad (5.14)$$

where η is the percentage (expressed as a fraction) of the maximum excitation used and v is the maximum heat added to the system. The approximate control signal fed to the elemental filters was a Gaussian process with a mean of $u_{\text{true}}(t_i)$ and a standard deviation of $0.1u_{\text{true}}(t_i)$; see Figure 5.68 for a graphical depiction of the true heating profile (solid black line) plus two realizations of the approximate profile (given by the dashed and dash-dot lines). Additionally, the relatively low melting point of aluminum (660 °C, see [1]) compared to the other materials limits the amount of (excitation) heat that we can safely apply to the system since we *did not know* what the material was beforehand.

We shall report the findings for four experiments that differ only in the level of system excitation used to aid the system identification process. The first experiment under the Simulation 6 heading sought to identify the thermal diffusivity constant (the unknown system parameter) using a “finely” discretized bank of elemental filters in the presence of abundant system excitation (limited so that we don’t melt the rod if it happens to be made of aluminum) to ascertain how quickly the system parameter could be identified. We excited the system by heating the left end ($\rho = 0$) of the rod for the first 0.75 seconds of the one-second duration run. [The accumulative effects

⁹We have assumed that our heating element is ideal and thus the temperature at the interface, i.e., at the left end of the rod, is perfectly regulated so that it matches the program given in Equation (5.14).

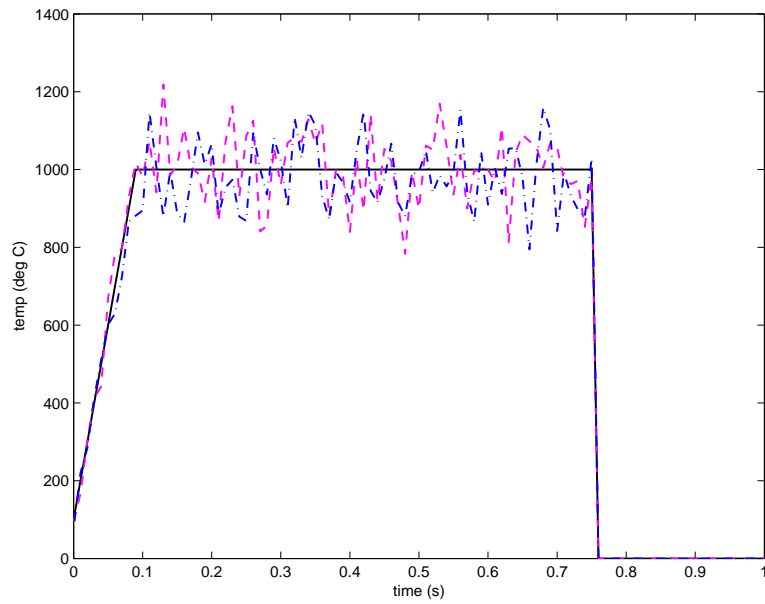


Figure 5.68 Simulation 6: The true rod heating profile (solid black line) applied to the left end of the rod, plus two realizations (dashed and dash-dot lines) supplied to the elemental filters for the maximum excitation case ($\eta = 1$) and temperature $v = 1000$ °C.

of the heating profile can be seen quite clearly in Figure 5.72(a) which shows the true and estimated temperature for the left end of the rod for the best modeled elemental filter; for 0.09 seconds the heat is steadily increased and is then held constant until 0.75 seconds.] Note that heating the rod is akin to dithering the actuator [57, 147, 186, 73] discussed in Section 2.4.11. By adding a known persistent excitation into the system, the identifiability of specific parameter of interest is increased. For the maximum level of persistent system excitation, the MMAE quickly identifies the thermal diffusivity constant as seen in Figure 5.69(a) [and Figure 5.72(h)]. Figure 5.69(a) shows clearly that elemental filter 3 is most likely correct, hence the material in question is aluminum with a probability of nearly one.

In a second experiment, we determined that even one fifth of the system excitation level used in the first experiment could achieve very good system identification results, i.e., we could still tell that the material in question is aluminum with about 90% certainty, as shown in Figure 5.69(b). At one tenth of the excitation, the prob-

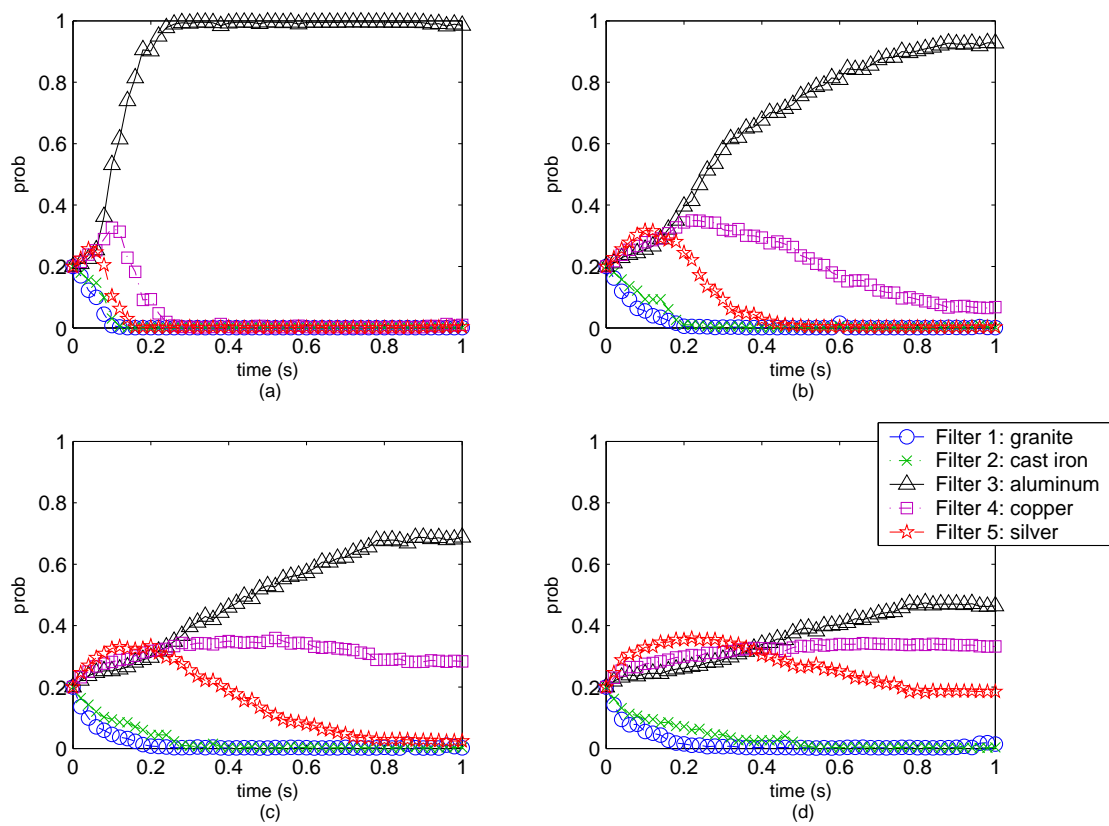


Figure 5.69 Simulation 6: Hypothesis conditional probability. (a) Maximum excitation, (b) Moderate excitation, (c) Low excitation, (d) Minimum excitation.

ability that the rod is aluminum drops to about 65%, with elemental filter 4 rising to almost one third probability that the material is copper; see Figure 5.69(c). At this low level of excitation, the three elemental filters with large diffusivity constant “look” and perform equally well in the beginning, the two filters with small κ ’s overestimate how hot the rod is and consequently have terrible residuals, and eventually the relatively high diffusivity constant filters received all of the probability. In the fourth case for one twentieth the original excitation magnitude, Figure 5.69(d) shows us that the elemental filter tuned for aluminum was still the most likely of the filters in the bank, with almost 0.5 probability, while the elemental filters featuring larger diffusivity constants (copper and silver) received about one half of the total probability. The full workup of plots are shown for this minimum level of excitation

in Figures 5.76 on page 5-174 through 5.81. The individual elemental filter plots associated with the moderate and low levels of excitation are not included in this report since the trends are all the same (as can be seen by comparing corresponding figures for the maximum and minimum system excitation cases).

When a filter is based on a model with a thermal diffusivity constant that is well below the true value, the filter “sees” a large increase in temperature as a tremendous increase in temperature since a “small” diffusivity constant means that the heat slowly diffuses through the material. Thus the huge errors indicated in Figure 5.70 are quite clearly seen in plots (a) through (f) and the RMS error values given in plots (i) through (p). While the absolute errors are much smaller when the system excitation is much lower, the same trends can be seen in Figure 5.76 for the elemental filter that is set up to identify granite.

At the other end of the spectrum, when the assumed κ is bigger than the true diffusivity constant, the filter underestimates the true level of heat flow because it “thinks” that the flow moves more swiftly than it truly can in the truth system with the smaller κ . Once again, the maximum excitation experiment provides the best example of this in Figure 5.74, and the minimum excitation plots exhibit a similar trend; however, the MMAE does not correct itself as quickly, as can be seen in Figure 5.80.

As expected, all of the elemental filters (whether properly modeled or not) saw their state estimation errors decrease after the excitation was removed. This is precisely why we apply persistent system excitation, to enable the MMAE to tell the difference between the elemental filters. Without persistent excitation, the system is able to dampen out heat flow such that most any model will match the heat flow characteristics encoded in the system dynamics model.

Thus, at substantially lower levels of persistent excitation, we must more coarsely discretize the set of values used to represent the diffusivity constant so that the individual filters appear more distinct. Preliminary simulations indicate

that if we forego persistent system excitation that (1) the discretization level needs to be such that two (or even three) orders of magnitude separate the parameter values used to specify the elemental filters, and (2) the MMAE takes longer to make a “decision”, therefore, the simulation time must be increased to observe the behavior of the filters.

There are of course an endless number of simulated experiments that we could perform in order to characterize the performance of the MMAE more fully for this problem. However, without actually running another simulation, we could hypothesize the following: how can we improve the simulation results for this minimal excitation experiment without coarsening the discretization? Recall that in the first simulation we demonstrated how the Q/R ratio influenced the distinguishability of the elemental filters. Since the diffusivity constant is part of the dynamics model, if we had conducted an experiment with a smaller Q/R ratio, i.e., the propagated estimates are to be favored over the measurement-updated estimates, then we could expect to enhance the results for the minimal excitation experiment because the differences between the filters would be accentuated.

Throughout this chapter, the likelihood quotient has proved to be an excellent guide to elemental filter performance since it considers both the residuals and the filter-computed residual covariance in concert with one another — see, for example, plots (g) of Figures 5.70 through 5.74 for one more set of illustrations. On the other hand, in this experiment, plots (b), (d), and (f) of Figures 5.70 through 5.74 also demonstrate clearly which elemental filter best represents the true system. A quick review of these plots shows that only the third elemental filter has a mean estimation error (given by the solid line) consistently bounded by the pair of dashed lines which are used to represent how well the filter “thinks” that it is performing. When the solid line (actual performance) radically differs from the expected performance given by the dashed lines, then we can be fairly certain that the model is not well matched to the real world that we are simulating. This same trend is fairly evident for the

minimal excitation case too, see plots (b), (d), and (f) of Figures 5.76 through 5.80, with the exception of filter 4 which looks nearly as well matched as filter 3 (to the untrained eye).

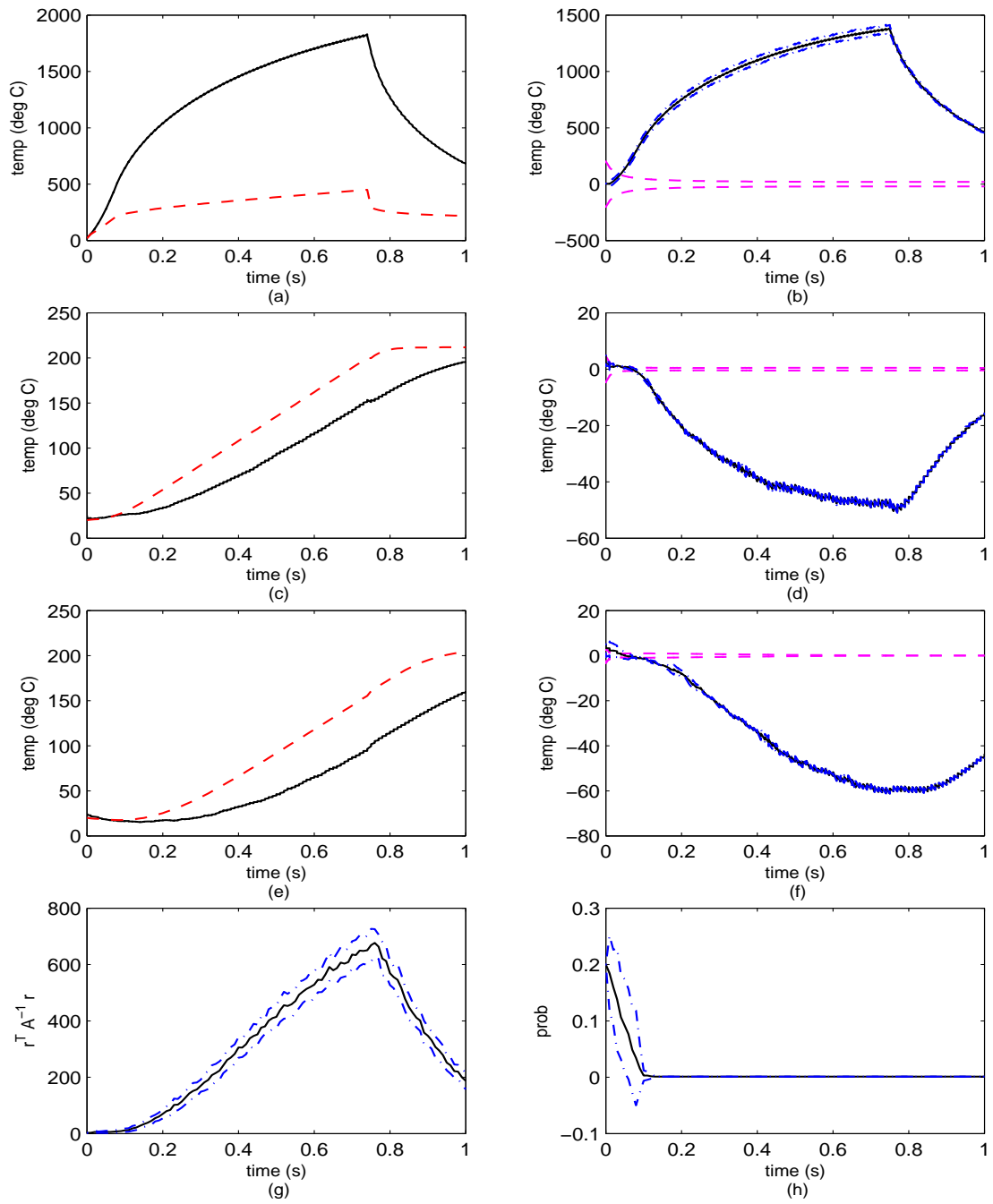


Figure 5.70 Simulation 6 (maximum excitation): Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

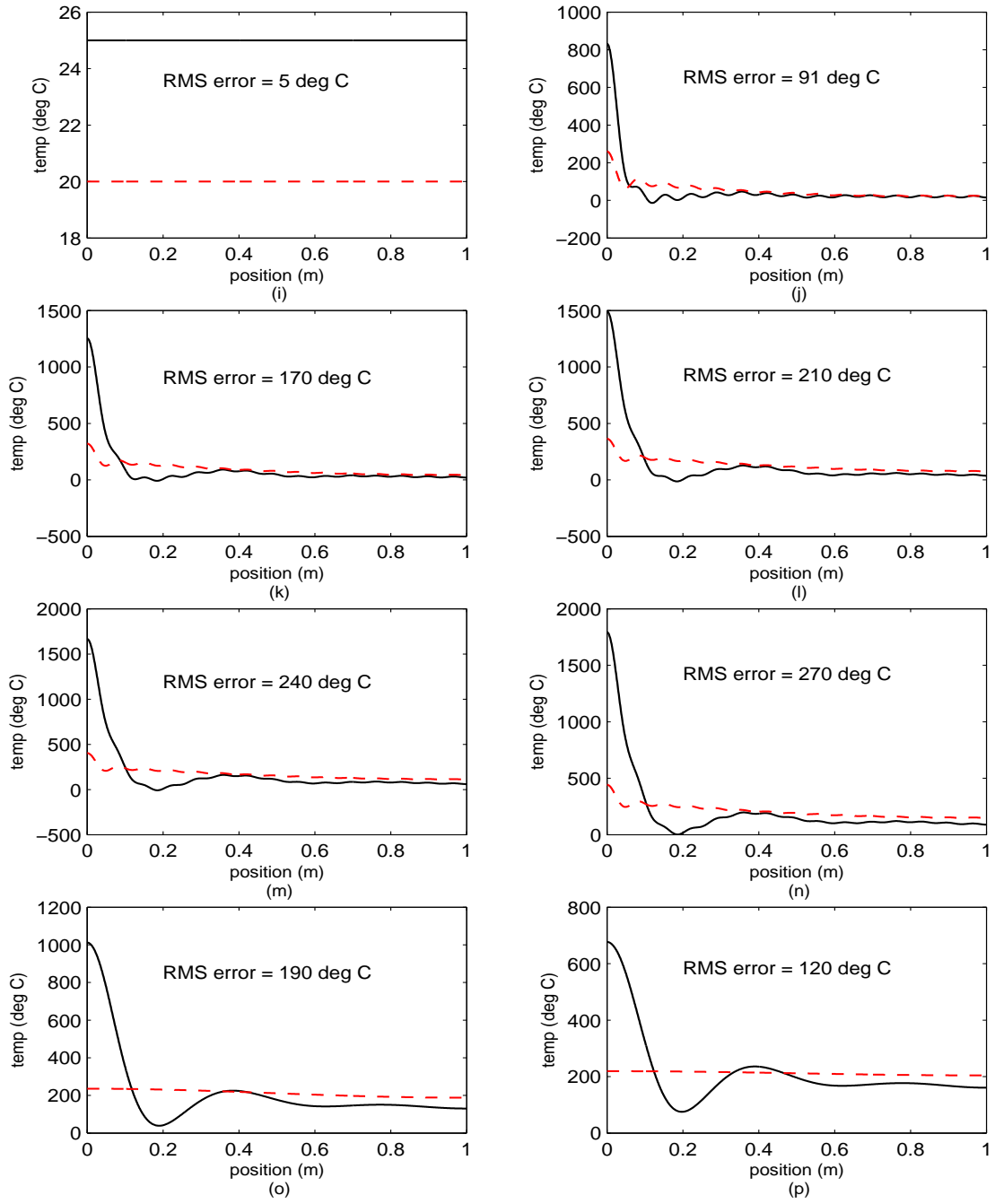


Figure 5.70 Simulation 6 (maximum excitation): Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

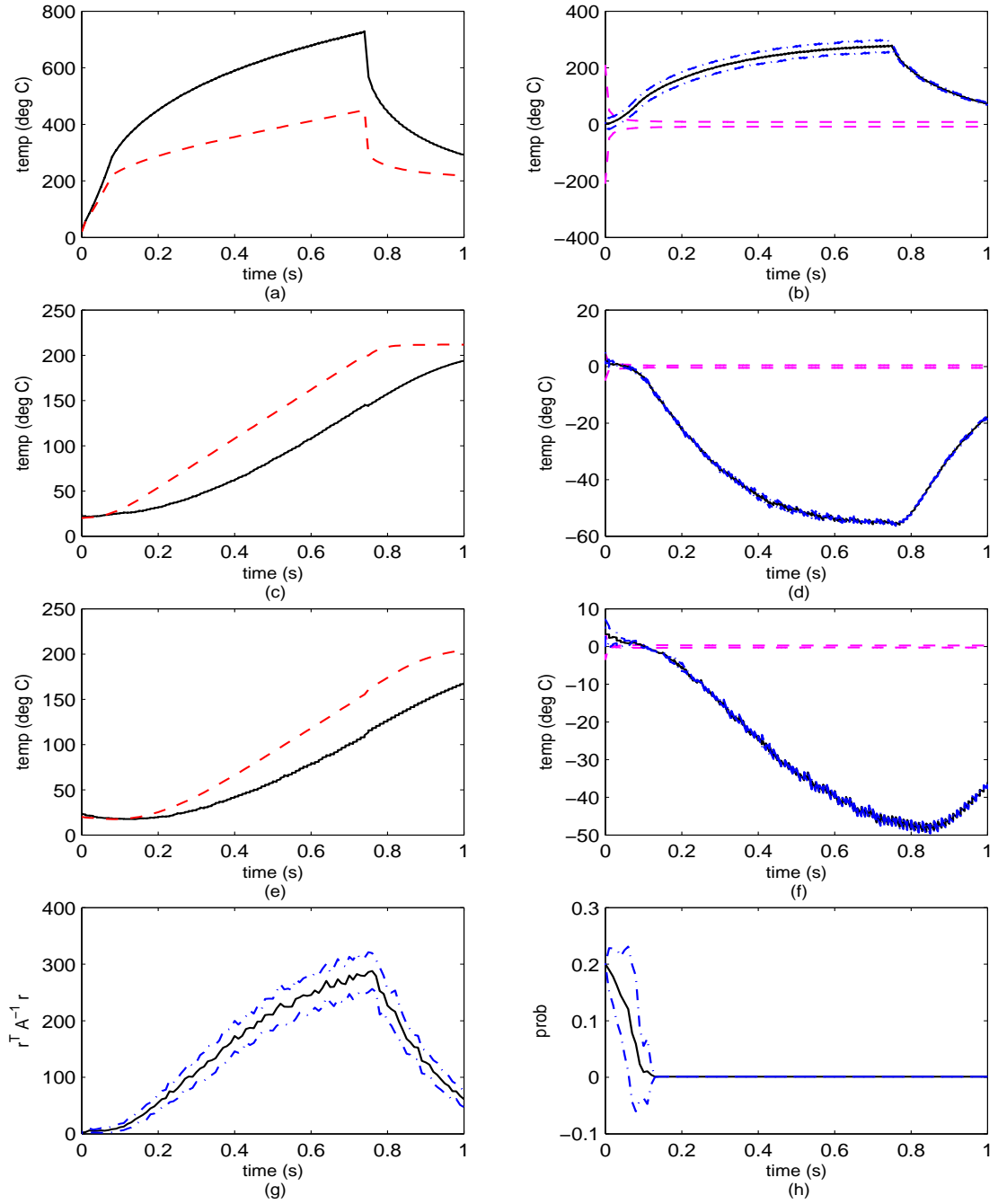


Figure 5.71 Simulation 6 (maximum excitation): Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

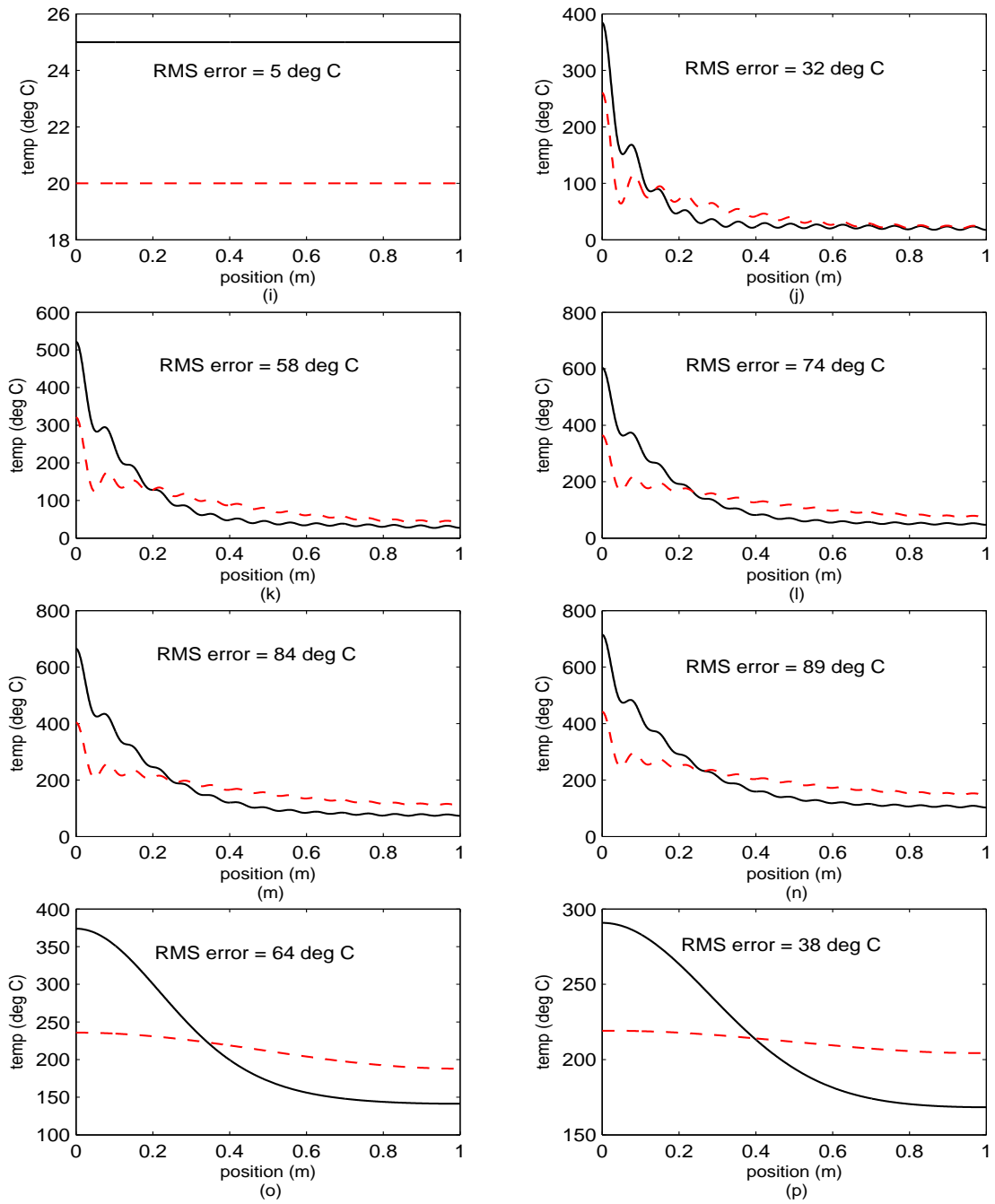


Figure 5.71 Simulation 6 (maximum excitation): Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

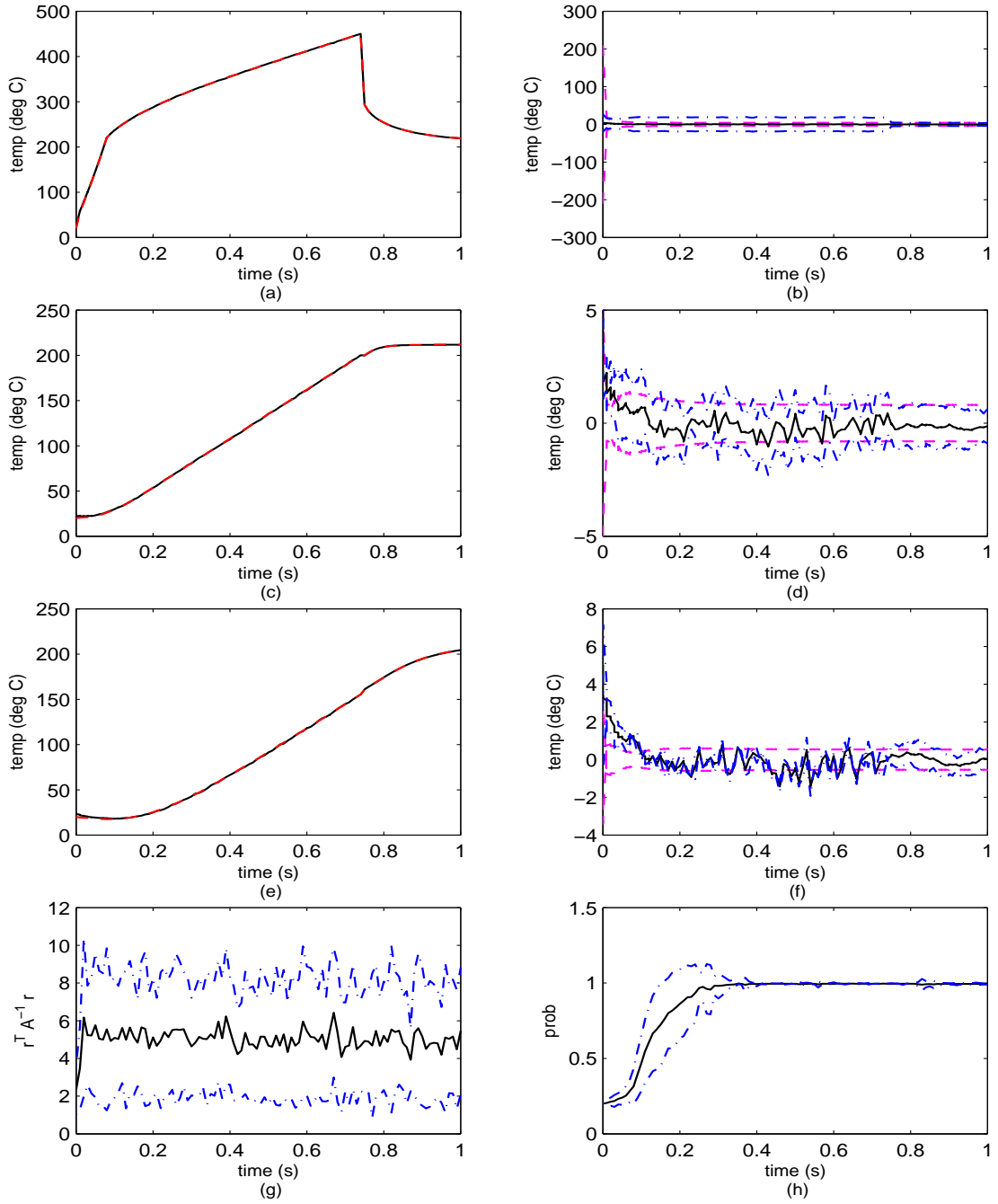


Figure 5.72 Simulation 6 (maximum excitation): Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

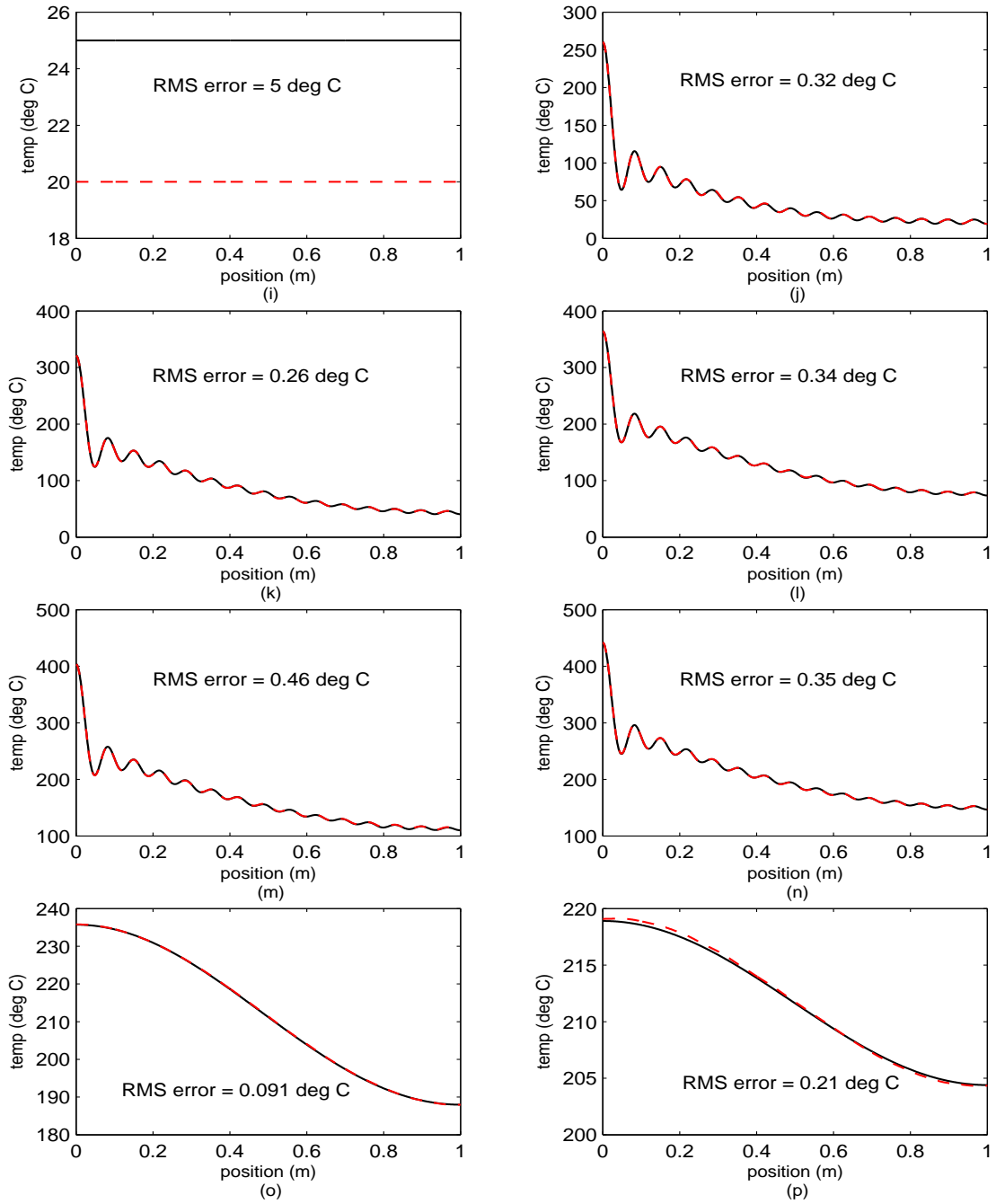


Figure 5.72 Simulation 6 (maximum excitation): Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

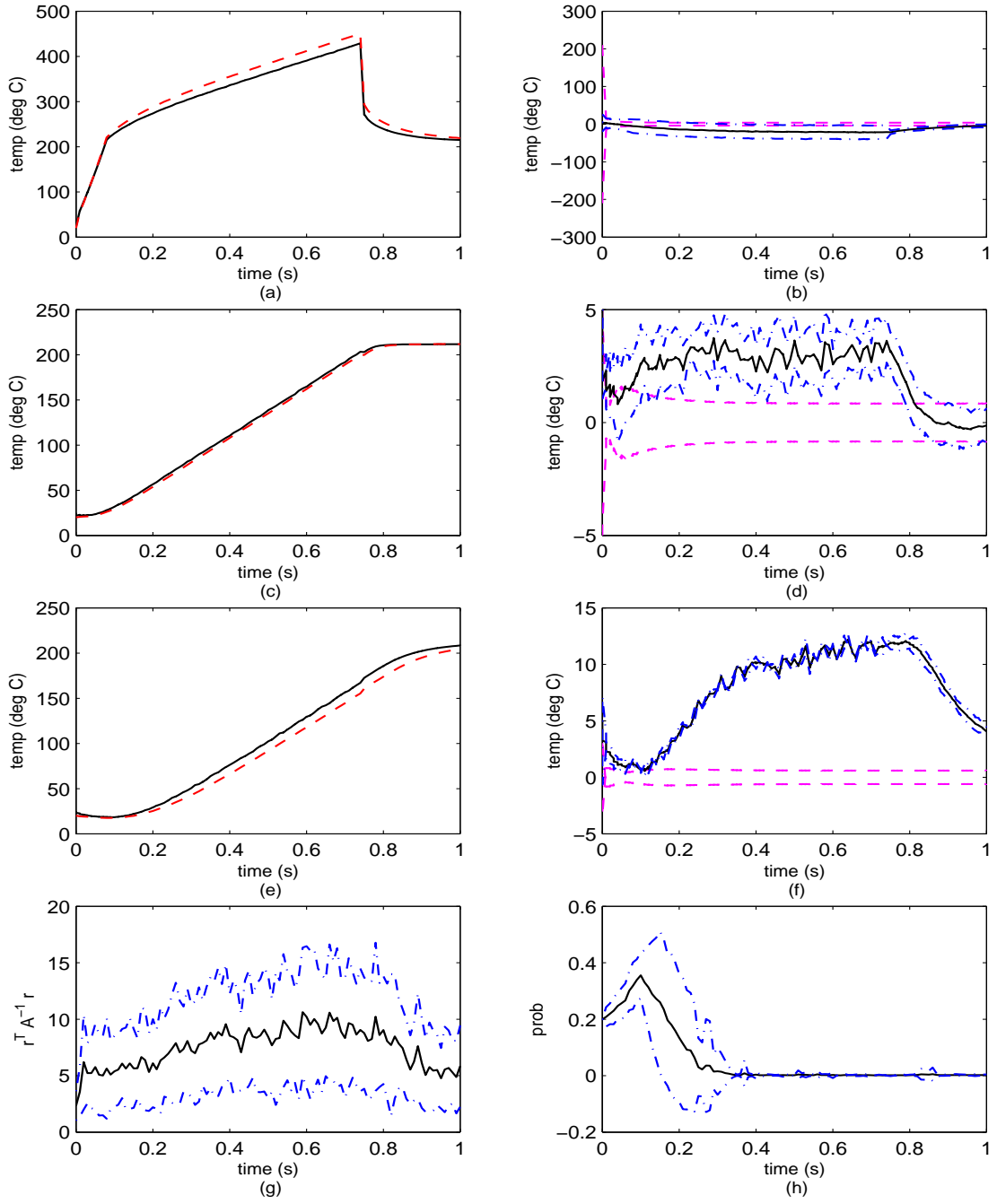


Figure 5.73 Simulation 6 (maximum excitation): Elemental Filter 4. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

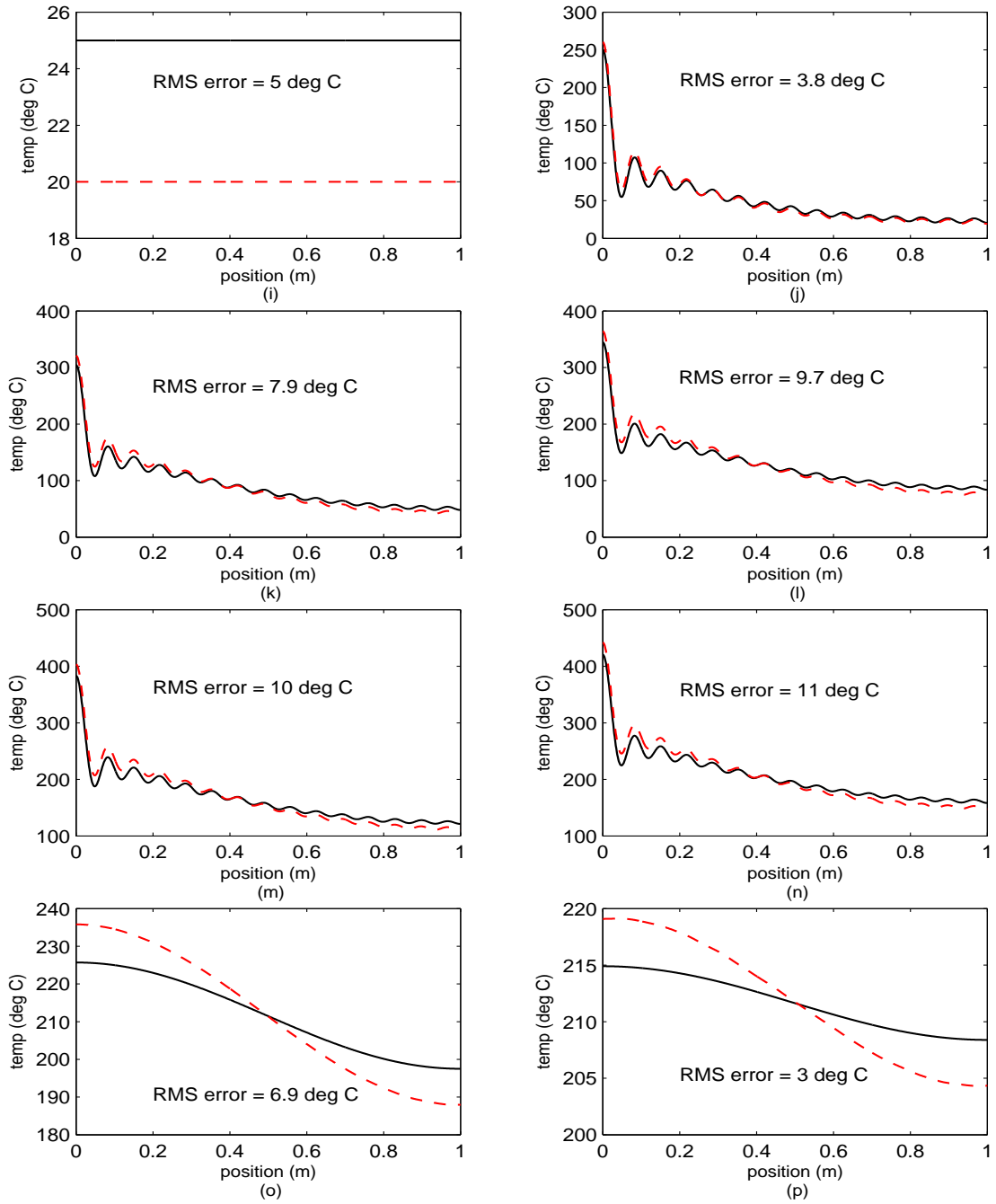


Figure 5.73 Simulation 6 (maximum excitation): Elemental Filter 4 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

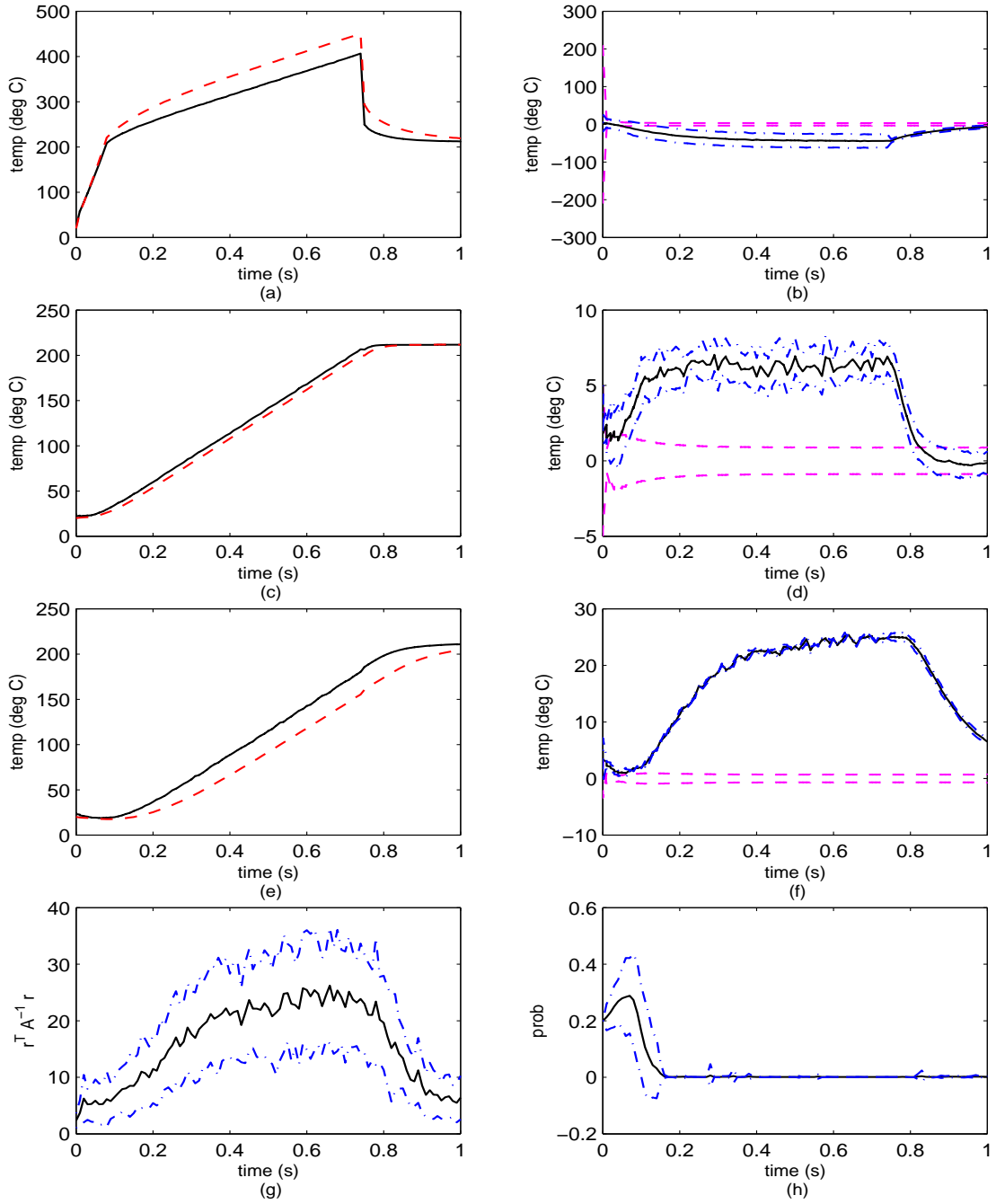


Figure 5.74 Simulation 6 (maximum excitation): Elemental Filter 5. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

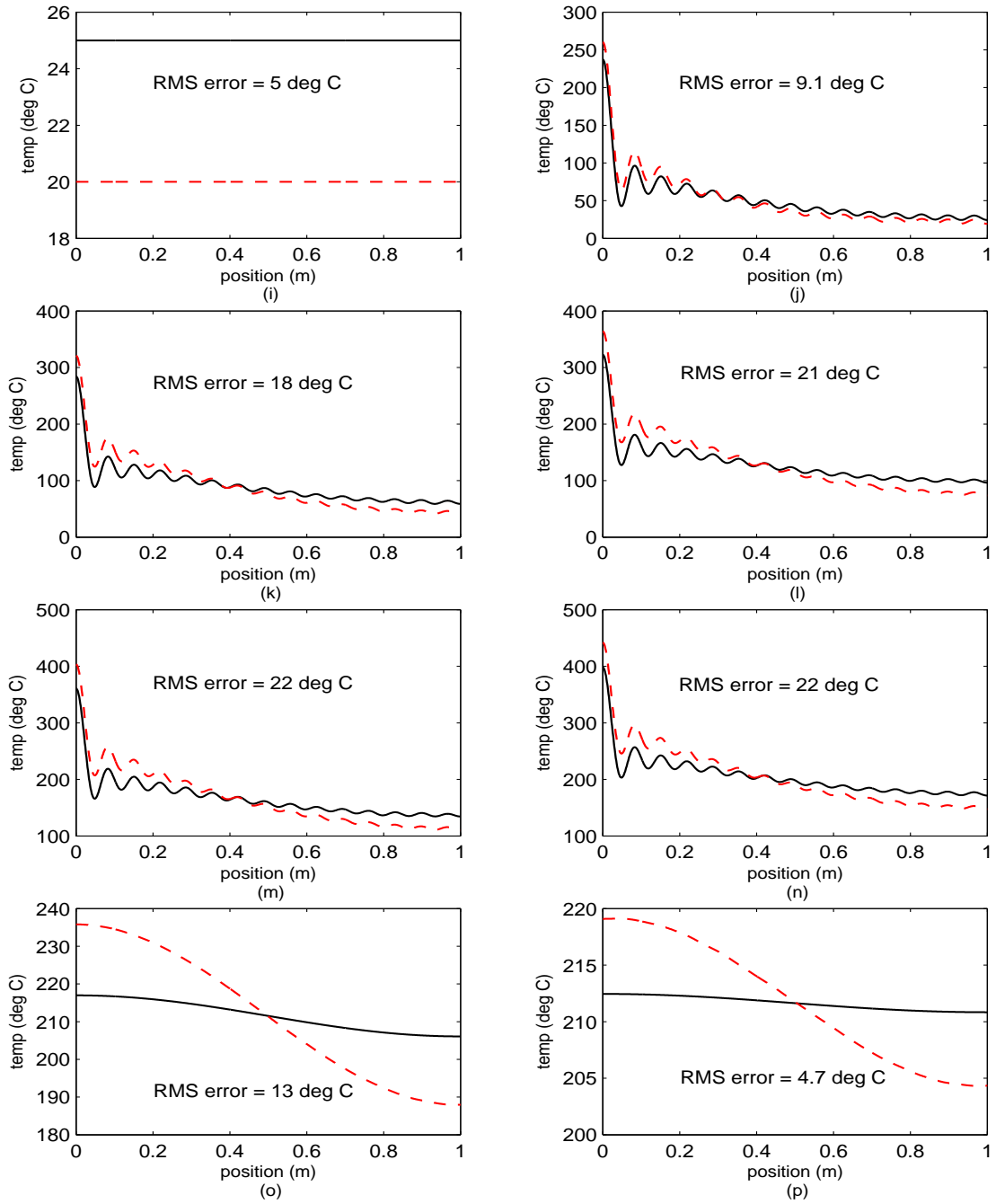


Figure 5.74 Simulation 6 (maximum excitation): Elemental Filter 5 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

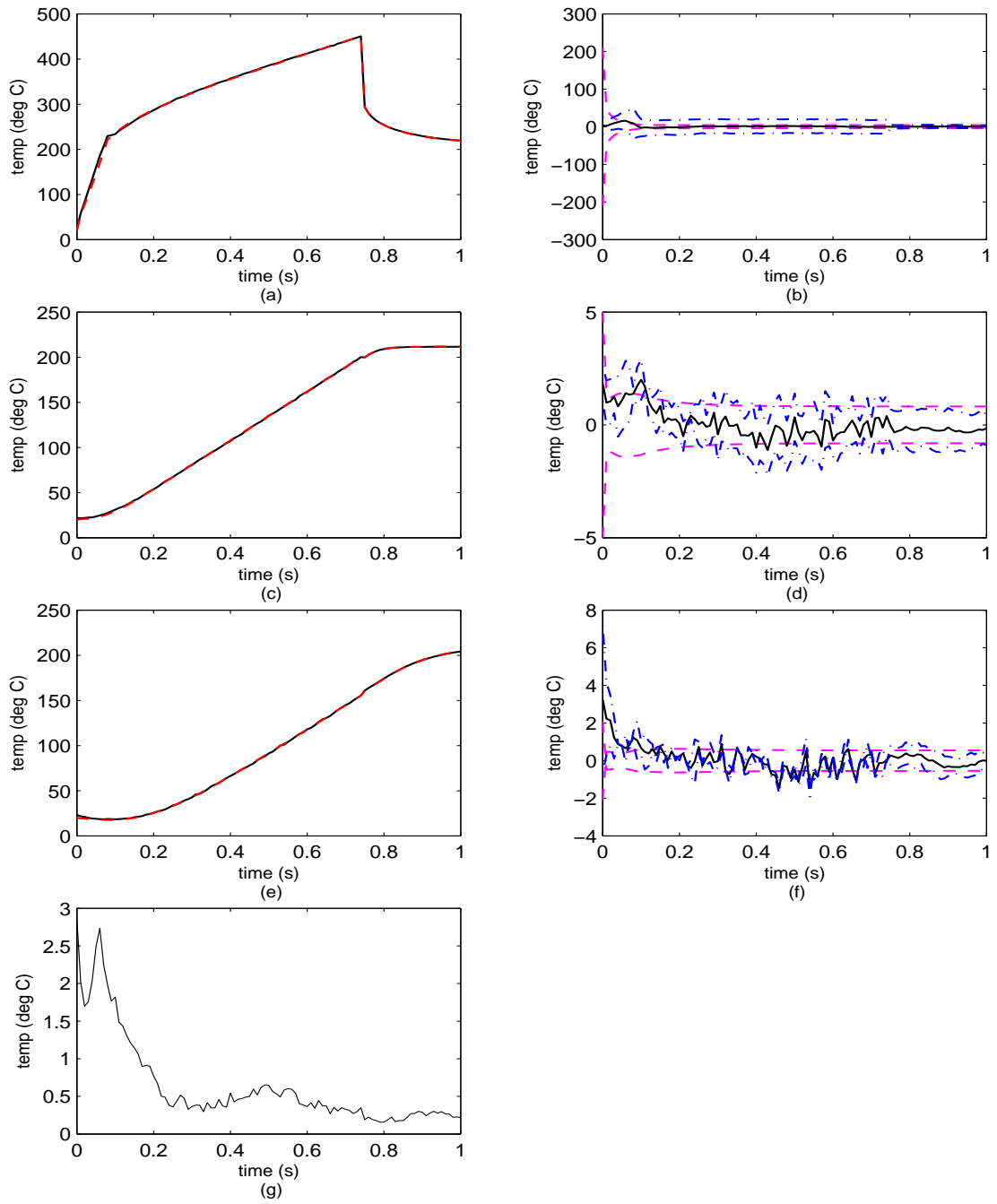


Figure 5.75 Simulation 6 (maximum excitation): Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

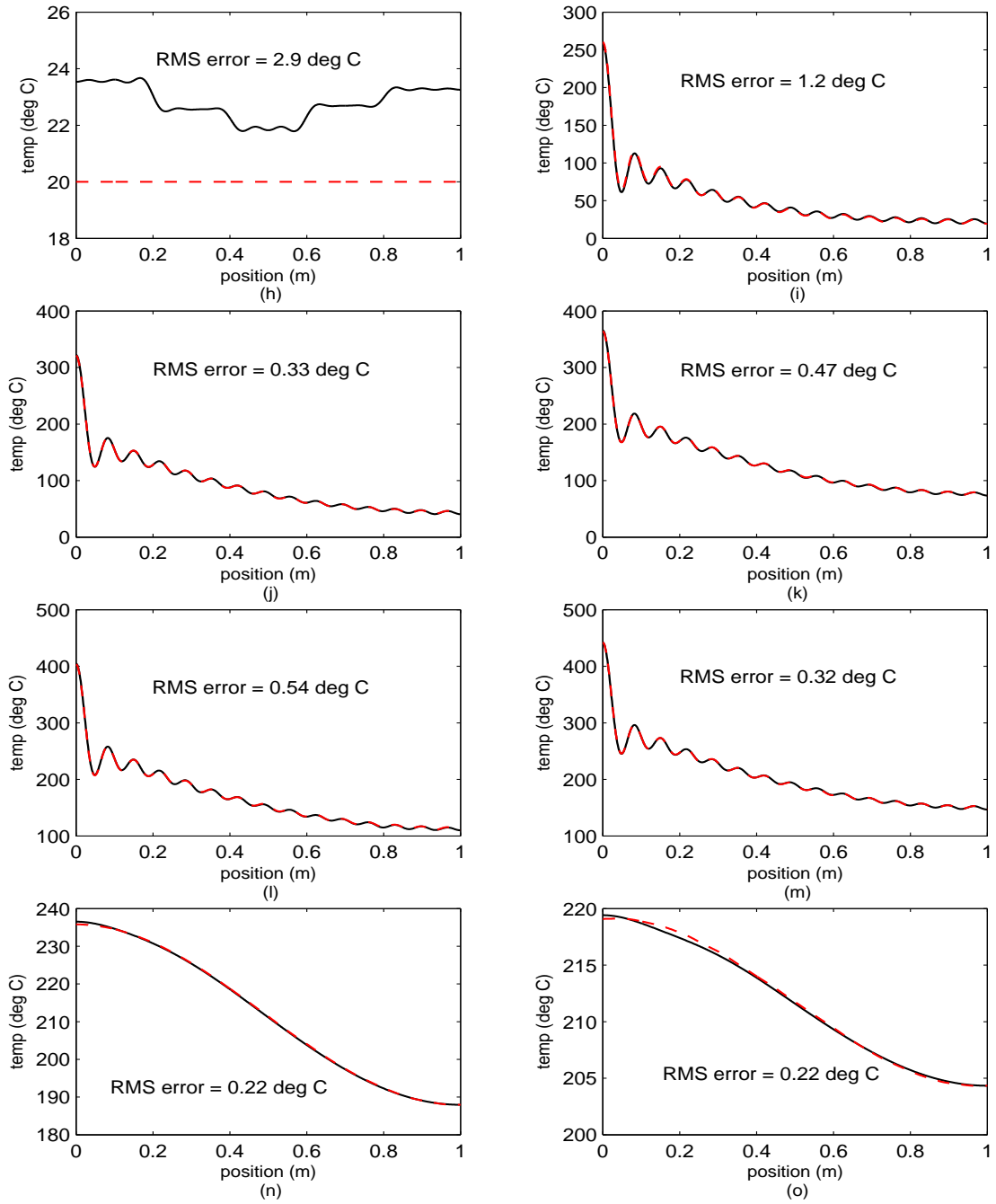


Figure 5.75 Simulation 6 (maximum excitation): Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

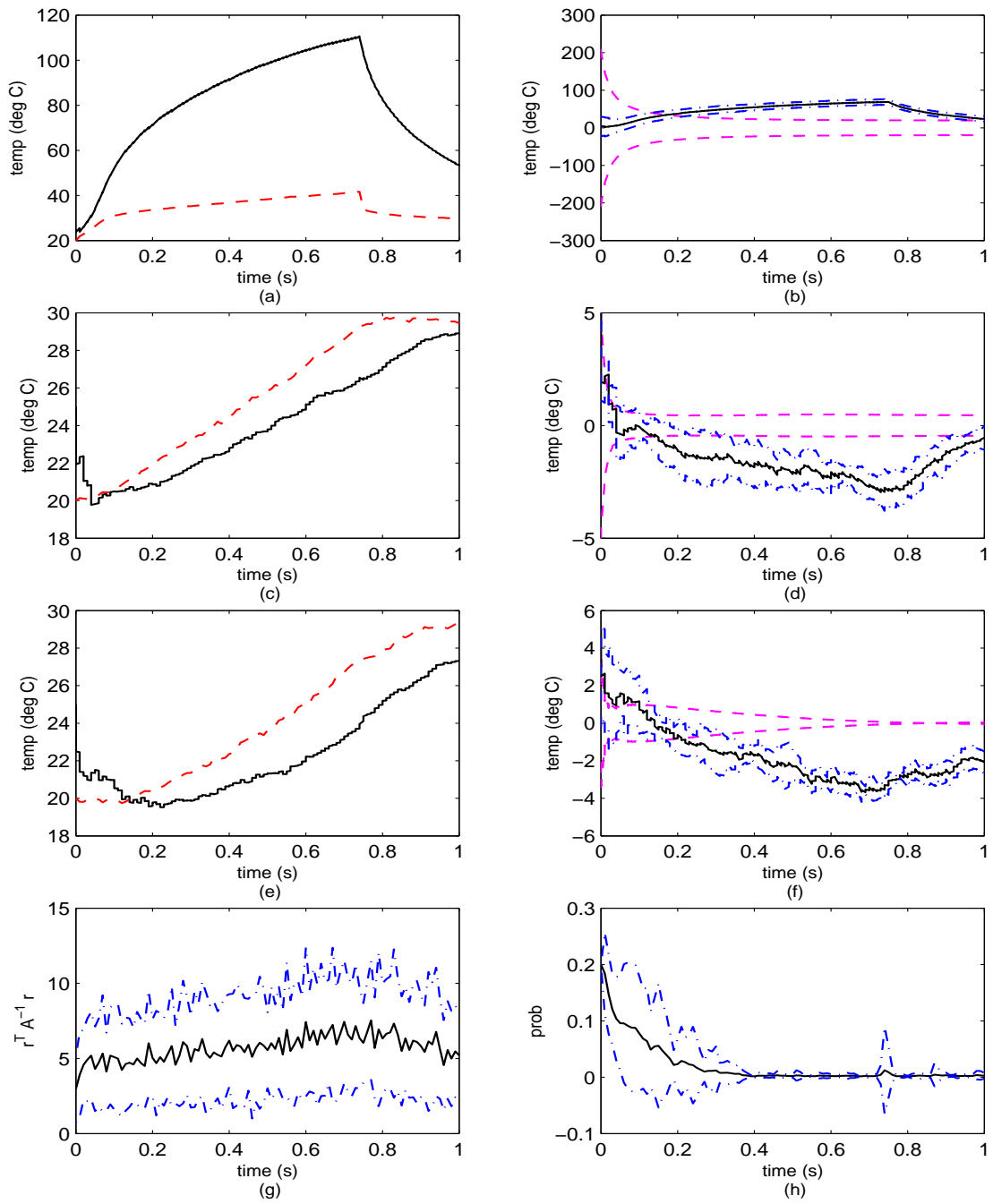


Figure 5.76 Simulation 6 (minimum excitation): Elemental Filter 1. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

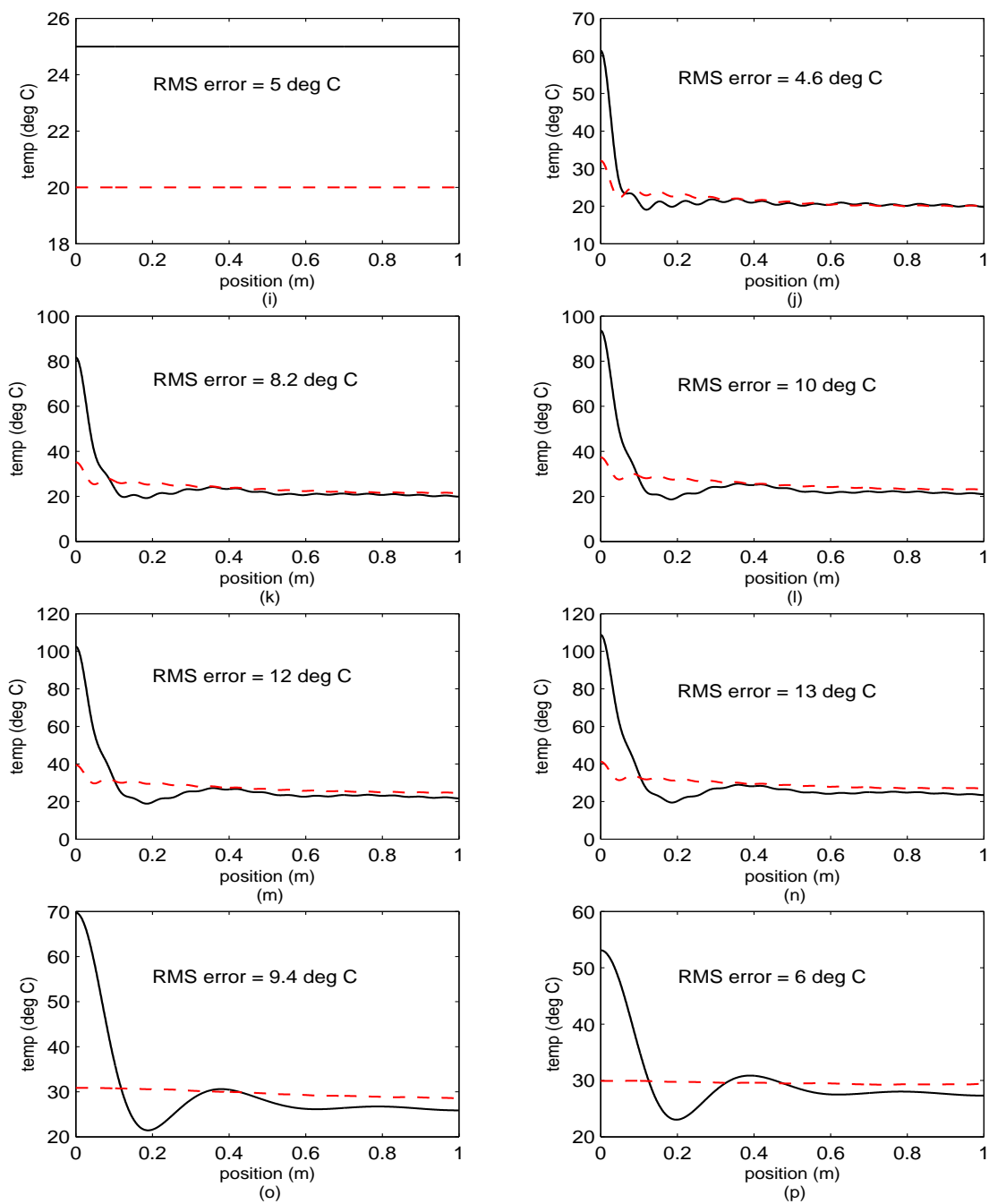


Figure 5.76 Simulation 6 (minimum excitation): Elemental Filter 1 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

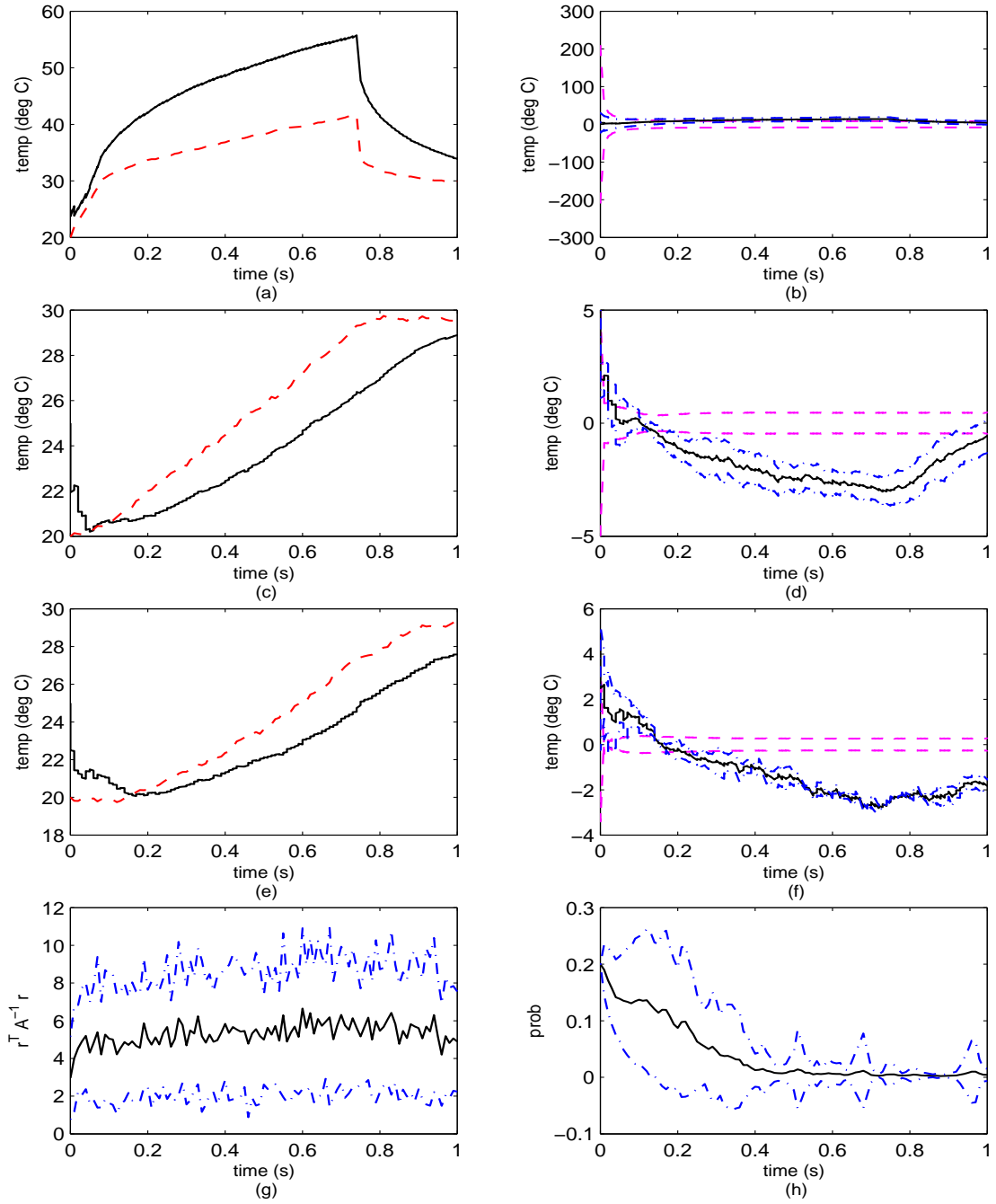


Figure 5.77 Simulation 6 (minimum excitation): Elemental Filter 2. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

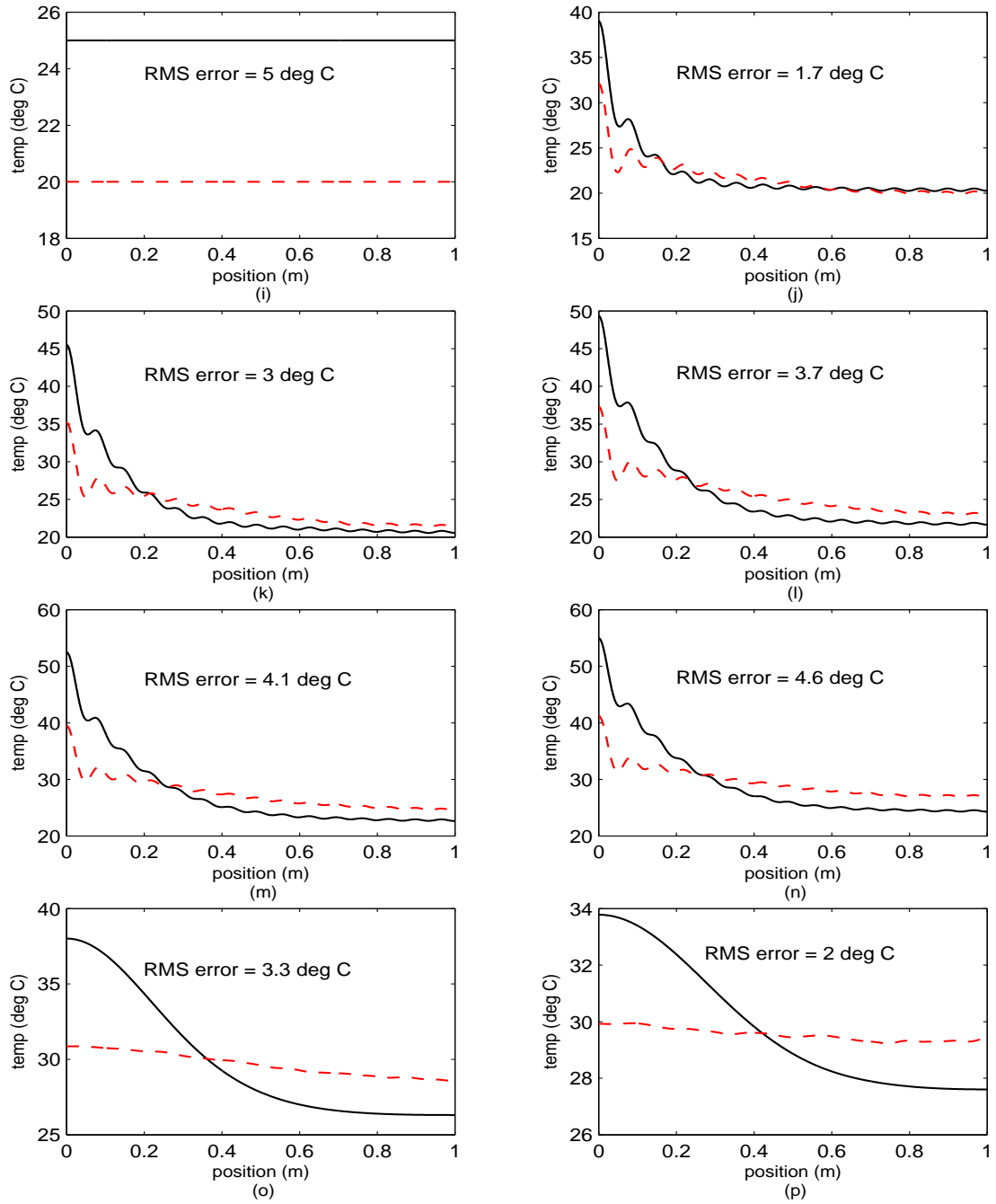


Figure 5.77 Simulation 6 (minimum excitation): Elemental Filter 2 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

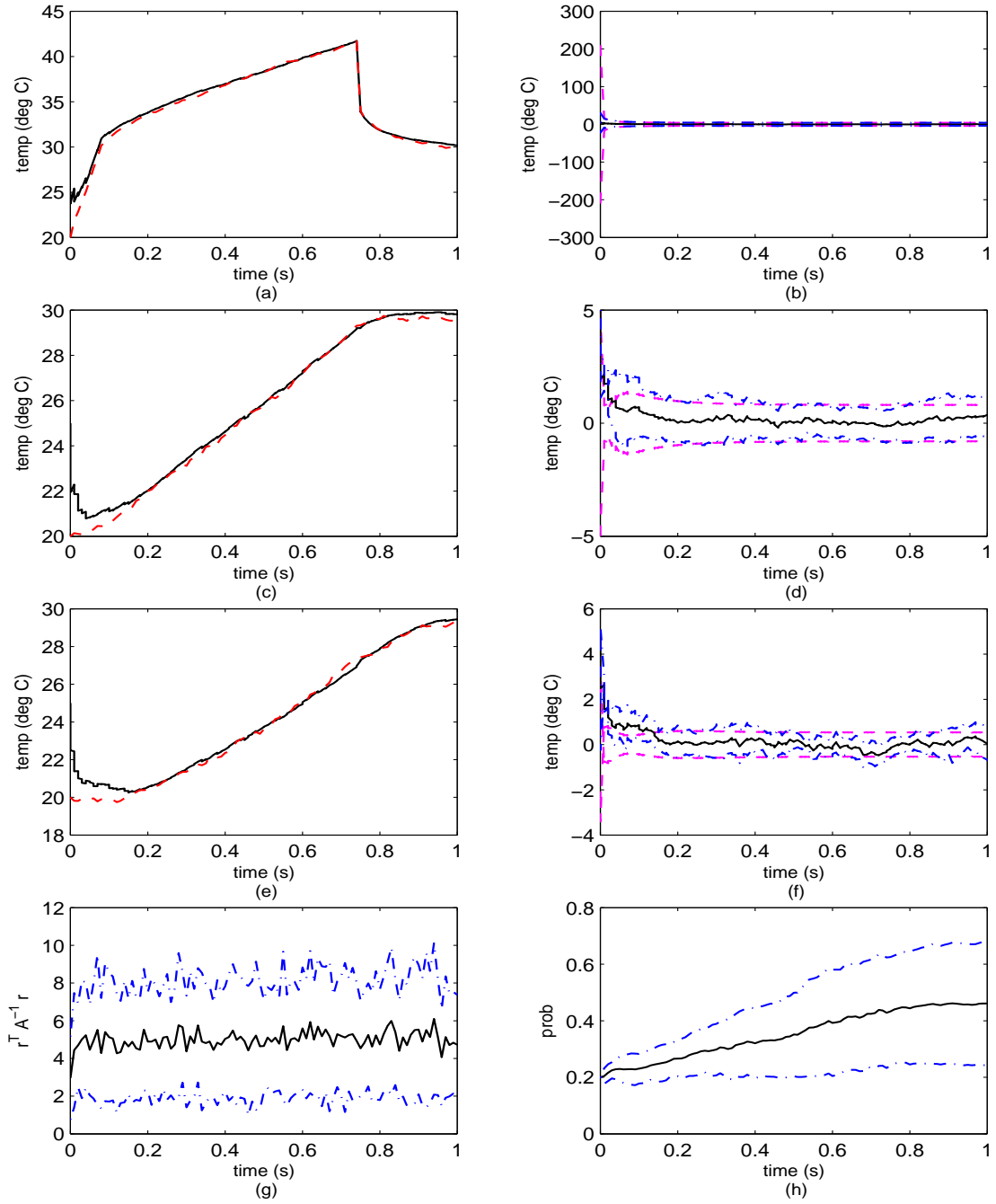


Figure 5.78 Simulation 6 (minimum excitation): Elemental Filter 3. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

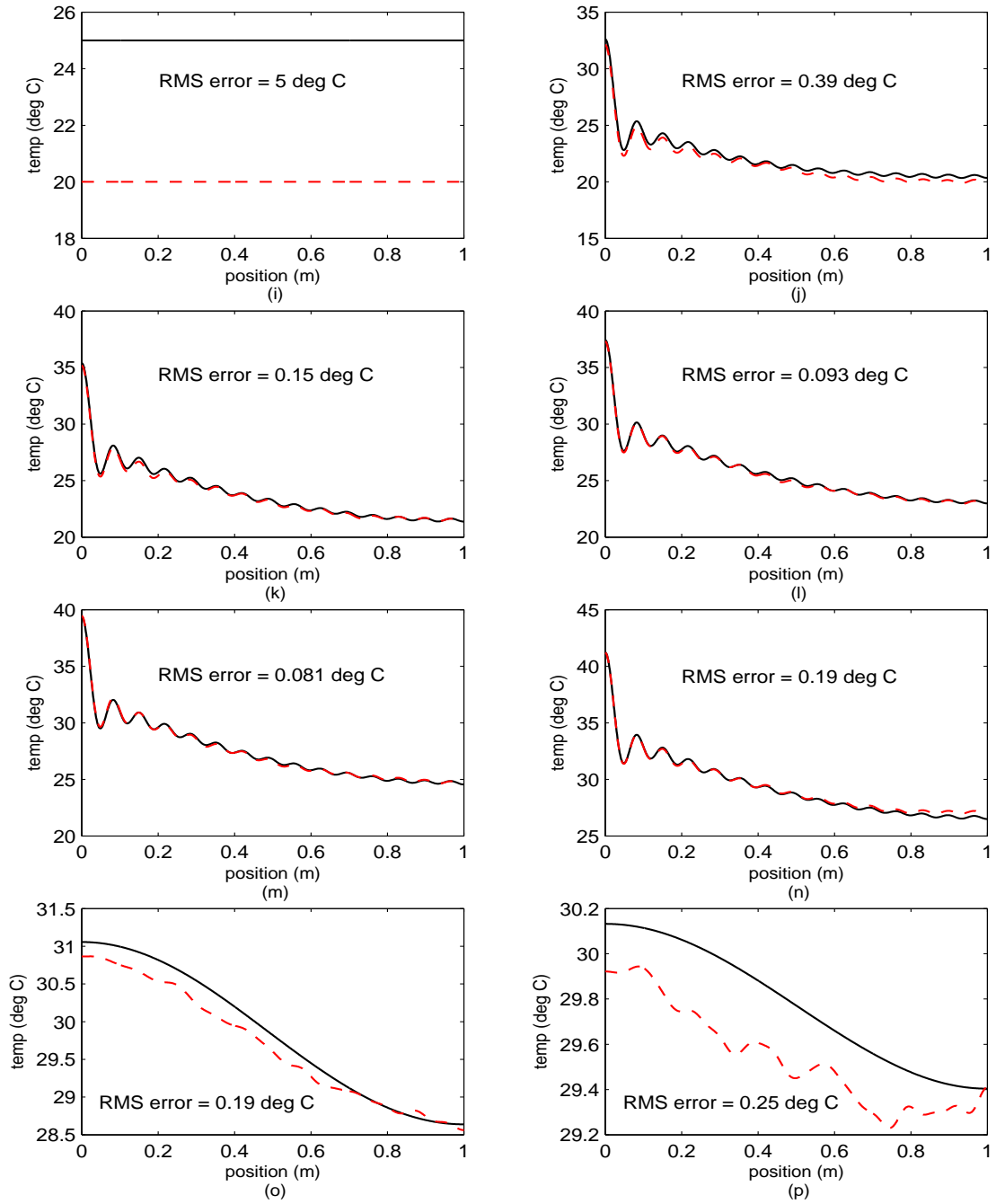


Figure 5.78 Simulation 6 (minimum excitation): Elemental Filter 3 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

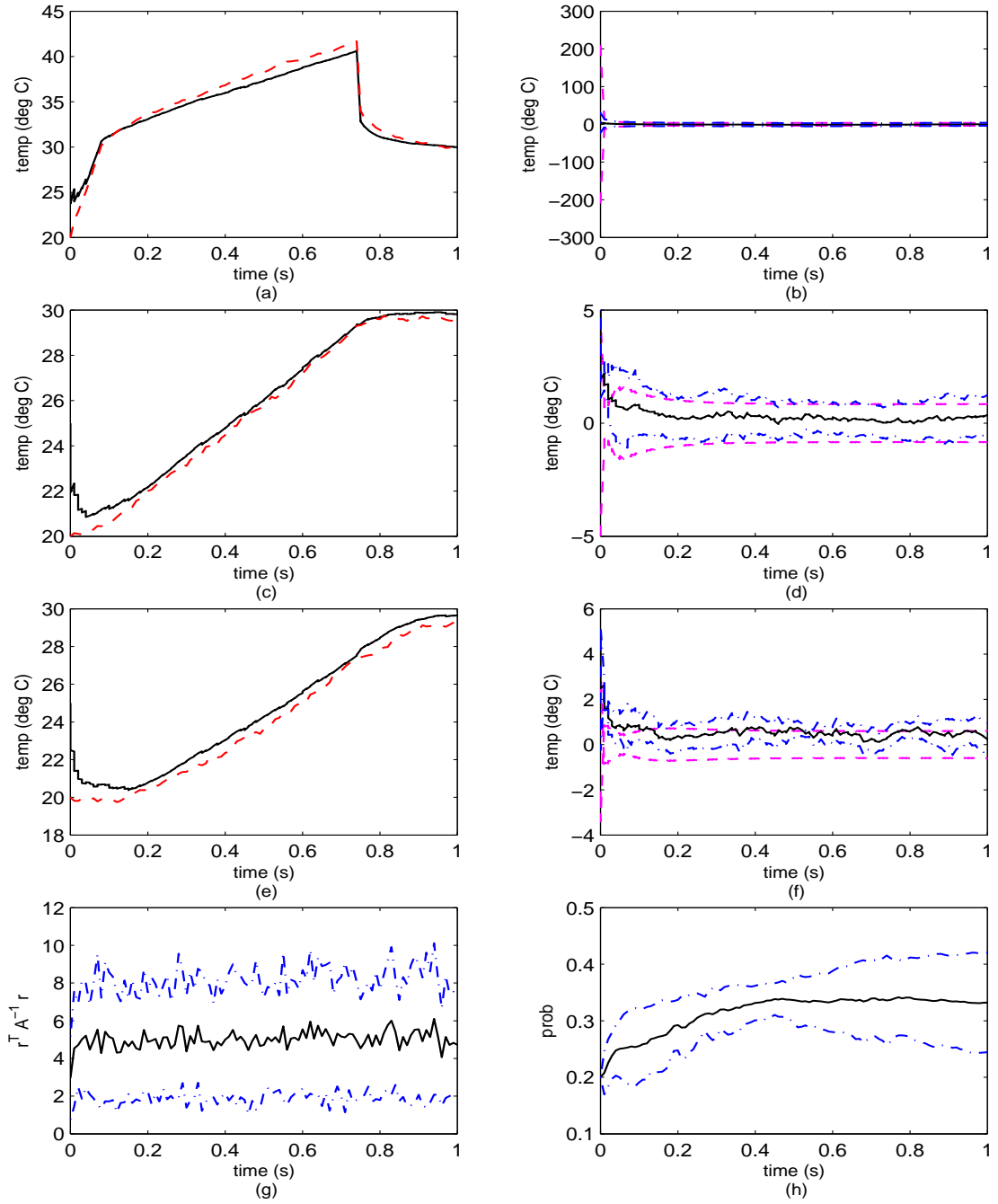


Figure 5.79 Simulation 6 (minimum excitation): Elemental Filter 4. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

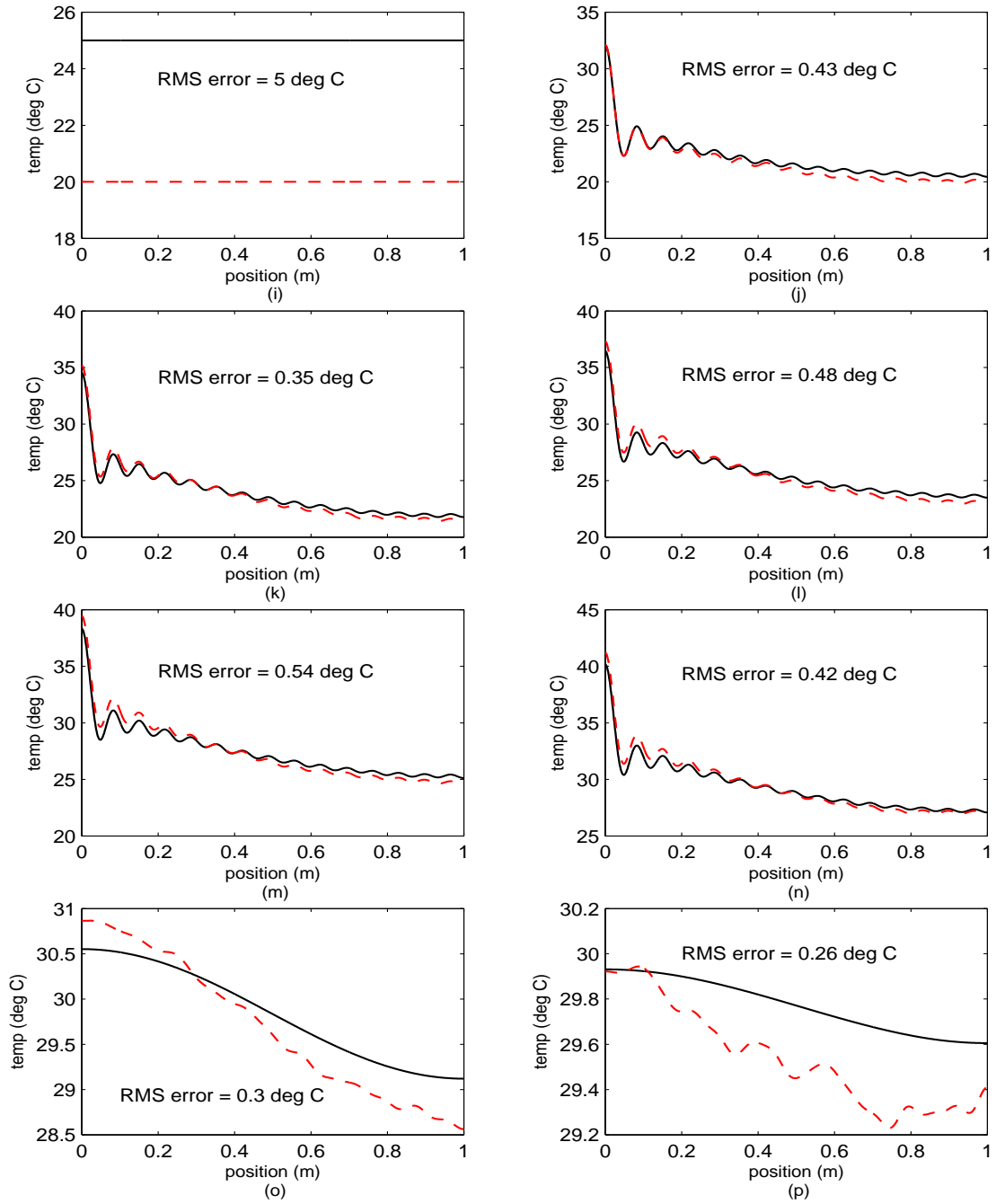


Figure 5.79 Simulation 6 (minimum excitation): Elemental Filter 4 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

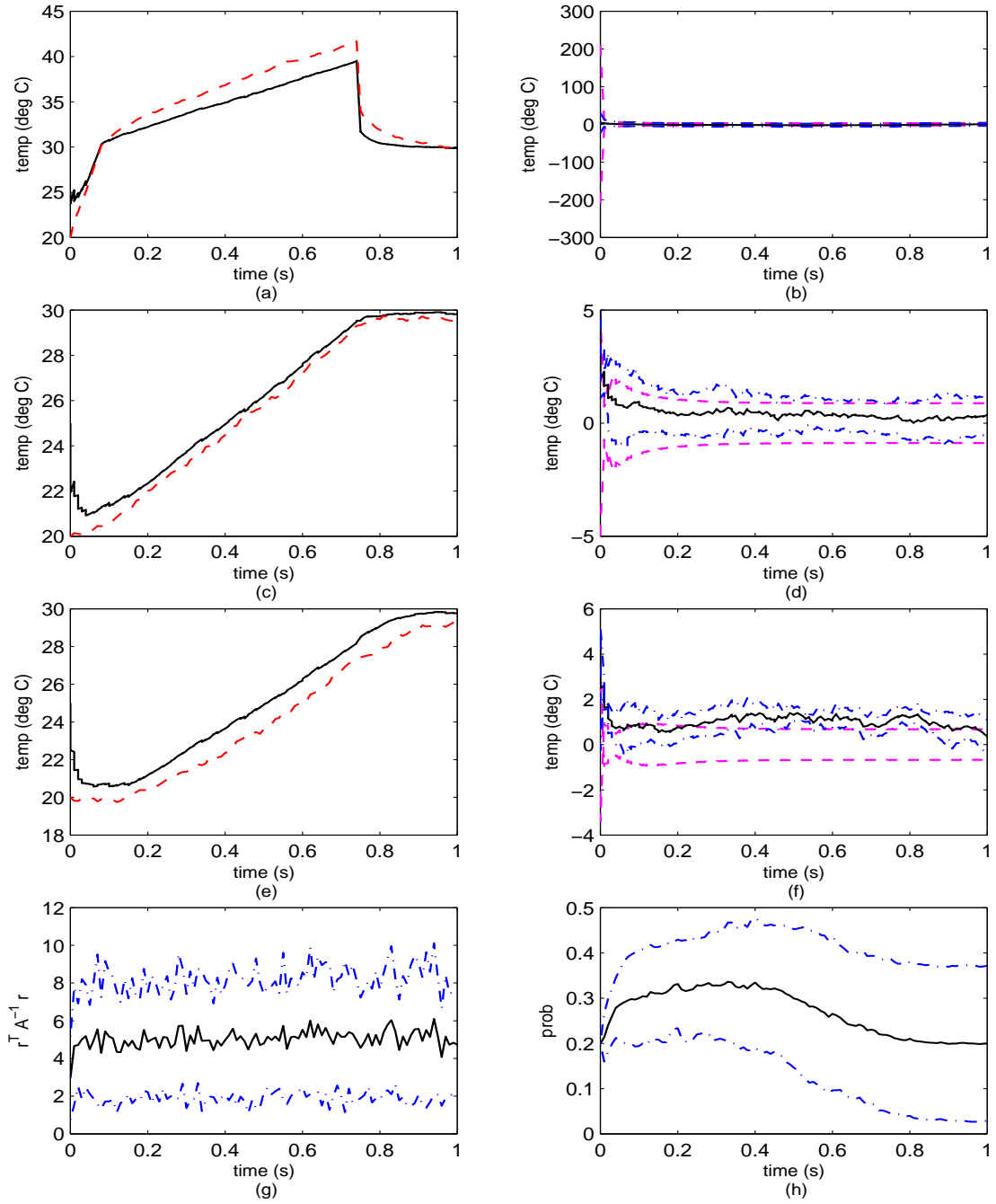


Figure 5.80 Simulation 6 (minimum excitation): Elemental Filter 5. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Likelihood quotient. (h) Hypothesis conditional probability.

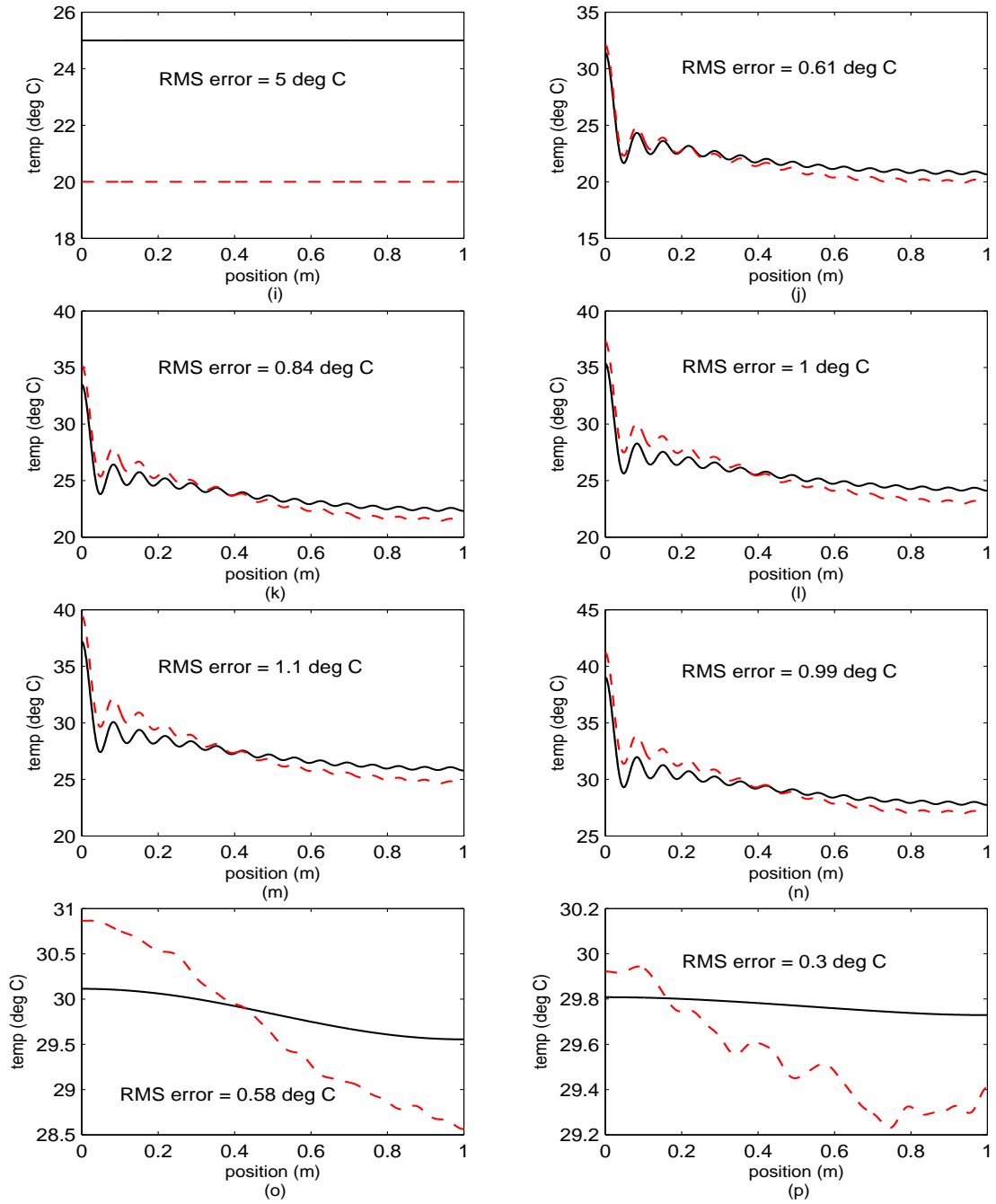


Figure 5.80 Simulation 6 (minimum excitation): Elemental Filter 5 (cont'd). (i) Rod temperature at $t_i = 0$ sec. (j) Rod temperature at $t_i = 0.14$ sec. (k) Rod temperature at $t_i = 0.29$ sec. (l) Rod temperature at $t_i = 0.43$ sec. (m) Rod temperature at $t_i = 0.57$ sec. (n) Rod temperature at $t_i = 0.71$ sec. (o) Rod temperature at $t_i = 0.86$ sec. (p) Rod temperature at $t_i = 1.00$ sec.

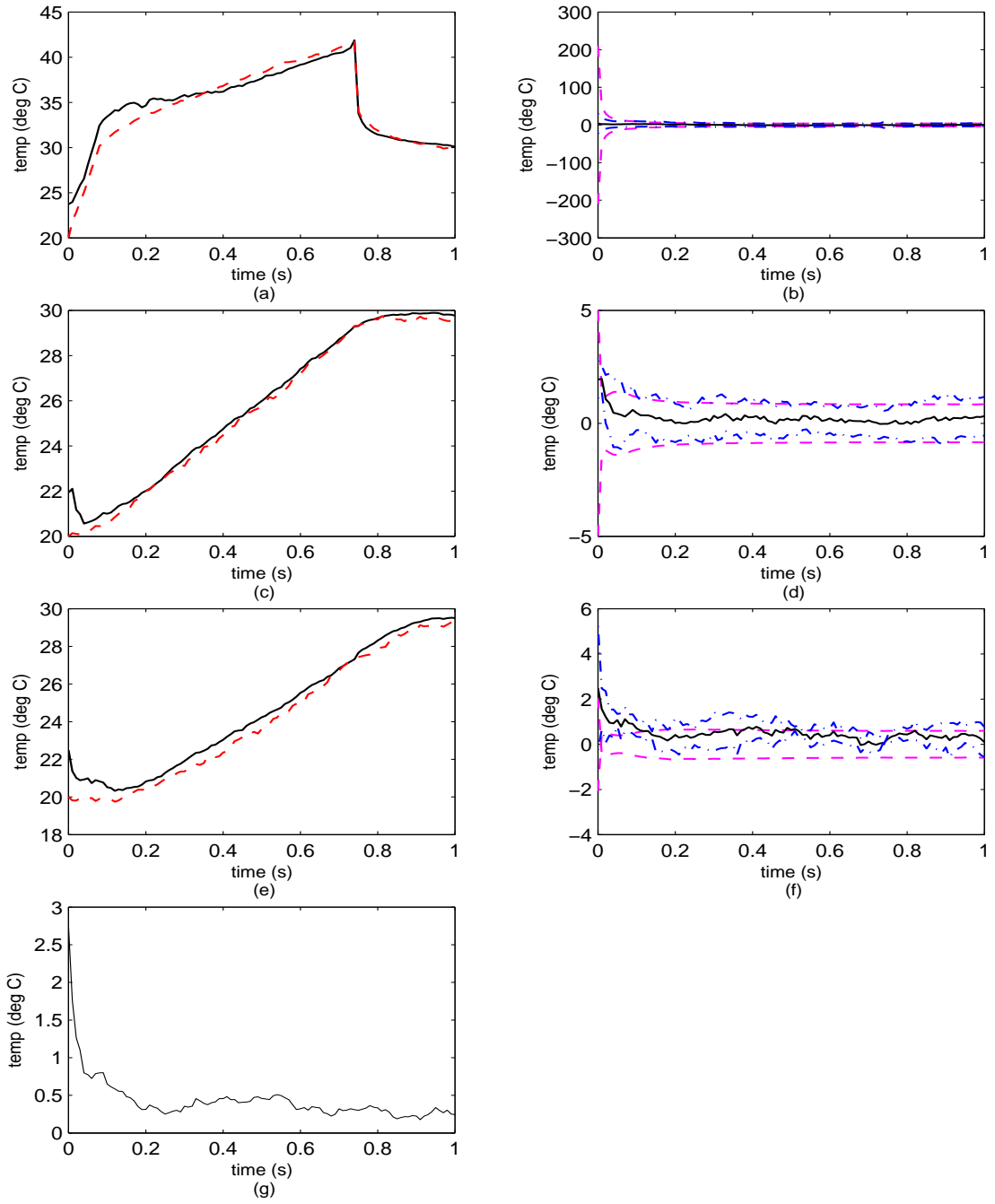


Figure 5.81 Simulation 6 (minimum excitation): Blended Filter. (a) Rod temperature at $\rho = 0$ m. (b) Error at $\rho = 0$ m. (c) Rod temperature at $\rho = 0.5$ m. (d) Error at $\rho = 0.5$ m. (e) Rod temperature at $\rho = 1$ m. (f) Error at $\rho = 1$ m. (g) Rod RMS temperature error.

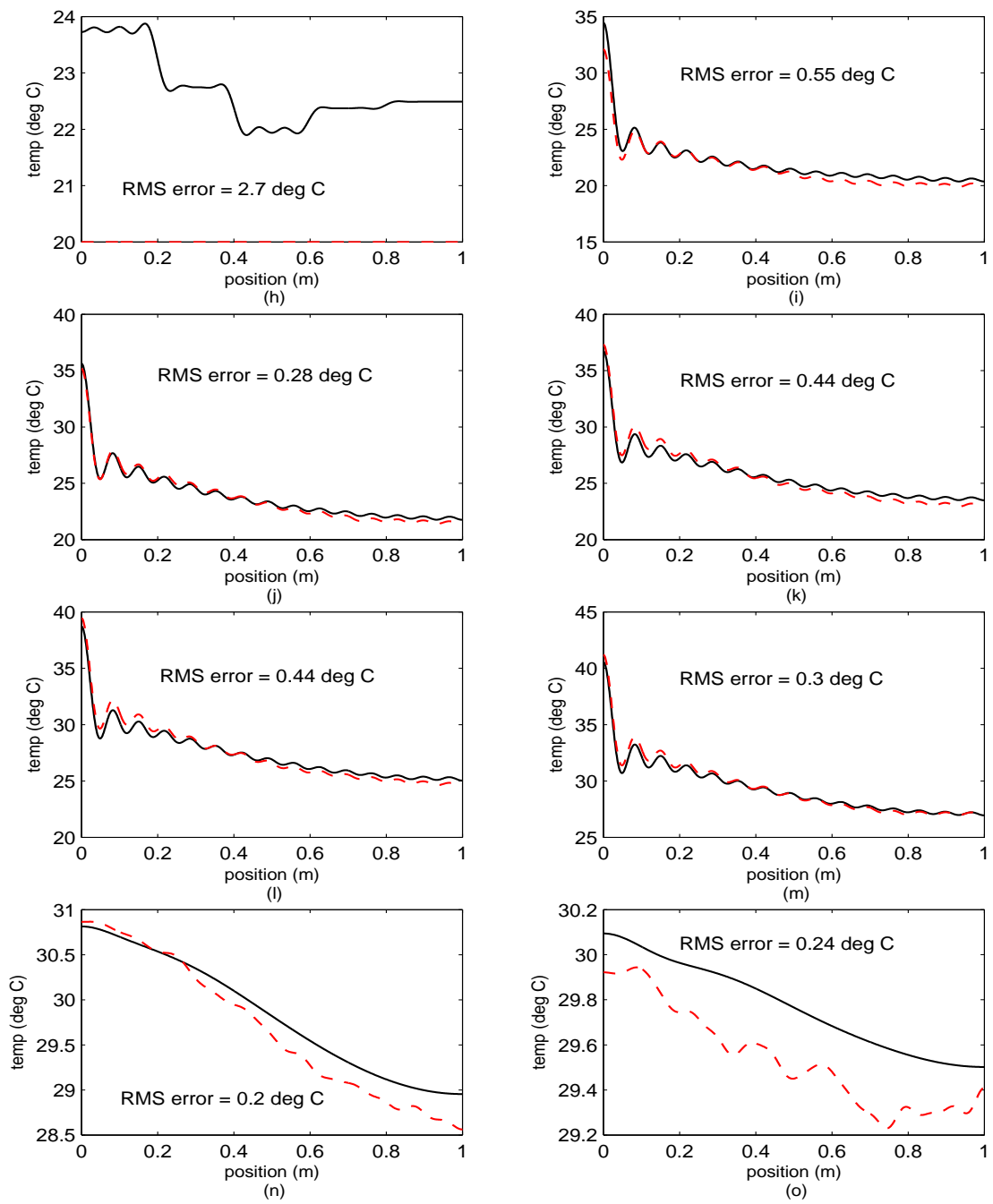


Figure 5.81 Simulation 6 (minimum excitation): Blended Filter (cont'd). (h) Rod temperature at $t_i = 0$ sec. (i) Rod temperature at $t_i = 0.14$ sec. (j) Rod temperature at $t_i = 0.29$ sec. (k) Rod temperature at $t_i = 0.43$ sec. (l) Rod temperature at $t_i = 0.57$ sec. (m) Rod temperature at $t_i = 0.71$ sec. (n) Rod temperature at $t_i = 0.86$ sec. (o) Rod temperature at $t_i = 1.00$ sec.

5.8 *Summary*

In the first five Monte Carlo simulations, we concentrated on adapting to an uncertain noise environment. Specifically, in the first simulation we concentrated on uncertain dynamics noise strength and showing how the Q/R ratio and discretization level, d , affected the distinguishability of the elemental filters in the filter bank as characterized by the hypothesis conditional probability histories. In the second simulation, we illustrated the importance of properly initializing the elemental filters. Simulations three through five featured the identification and adaptation of the MMAE to an unknown (constant, linearly varying, and abruptly changing) measurement-corruption noise covariance. Finally, in simulation six, we demonstrated perhaps the most powerful aspect of the MMAE, that of accurately identifying a structural aspect (or parameter) of the dynamical system, in this case, the thermal diffusivity constant.

VI. Conclusions

6.1 Introduction

The goal of this chapter is twofold: firstly, to summarize the contributions, both large and small, resulting from the research documented in this dissertation, and secondly, to suggest some recommendations for future research. The focal point of this dissertation is the infinite-dimensional sampled-data Kalman filter (ISKF) given by Theorem 91, on page 3-85. In general, the other contributions discussed below serve to motivate or support the development of the ISKF and the generalized infinite-dimensional multiple model adaptive estimation (GIMMAE) framework, or to illustrate their use in a practical problem. The recommendations for future work represent just a few of the interesting paths that should be investigated to develop this area of research more fully.

As illustrated in Table 3.1 on page 3-83, the Kalman filtering quartet, that began with the (discrete-time dynamics, sampled-data measurements) Kalman filter [95], the (continuous-time dynamics, continuous-time measurements) Kalman-Bucy filter [96], and the infinite-dimensional (continuous-time dynamics, continuous-time measurements) Kalman-Bucy filter by Falb [51] is now complete with the addition of the infinite-dimensional (discrete-time dynamics, sampled-data measurements) Kalman filter (ISKF). More importantly, the ISKF provides the proper foundation for crafting an exact and easily implemented algorithm for a digital computer, as demonstrated in Chapter IV.

6.2 Contributions

Our first contribution, which can be found in Section 1.3, is a simplified presentation of multiple model adaptive estimation (MMAE) and its application to navigation. This introductory development of the MMAE is intended to be a source

of motivation for new researchers — it was inspired by the illustrative “lost at sea” example given in Section 1.5 of Maybeck [129].

The focus of this dissertation is on the development of an abstract mathematical algorithm and then demonstrating how one might implement this algorithm for a practical problem. The background material given in Chapter II, while not normally considered a contribution, contains a large collection of previous contributions from the stochastic estimation literature; thus, it serves to not only prepare and motivate the reader for understanding this research, but to “prime the pump” for future research directions.

With an eye towards future research, the ISKF was built incrementally upon a sequence of stochastic estimators and increasingly specialized measurement models. While the derivations in this research were certainly inspired by the published works contained in the literature (to include the advanced topics discussed in Section 1.4, the extensive background leading to the present day MMAE in Chapter II, and linear estimation theory given in Section 3.3), only the first five definitions of Section 3.3 were borrowed (and cited) from the literature. The remainder of the development contained in Sections 3.3 through 3.6 represents an original contribution.

The development of the ISKF is underpinned by conditional expectation¹; thus, the conditional mean estimator posed on a separable Hilbert space, see Theorem 65, represents the core of the ISKF — our development of a conditional mean estimator parallels the finite-dimensional case given by Scharf [170]. Following the initial development, we showed that the conditional mean estimator solves the minimum mean-squared error estimation problem posed in Definition 68. After defining a measurement model for correlated states and observations (CSO) in Definition 70, a linear infinite-dimensional minimum variance unbiased estimator (LIMVUE) for CSO is given in Theorem 71; this theorem takes the first big step towards the ISKF. Next, a generalized linear measurement model is proposed in Definition 72 and then

¹See Definition 55 and the references therein for a general treatment of conditional expectation.

another LIMVUE is given based on this measurement model in Theorem 75. Now that the basic estimation theory for a state stochastic process is in place, we pose a generalized linear stochastic measurement model in Definition 76 and develop the corresponding LIMVUE for stochastic processes in Theorem 78. Finally, the ISKF is given by Theorem 91 and the GIMMAE framework is developed in Section 3.6.

While many (if not most) of the physical problems motivating this line of research mathematically model the system dynamics using an infinite-dimensional continuous-time description, such as a stochastic partial differential equation, our goal is to craft a digital computer algorithm, hence there is an intermediate need to transform the infinite-dimensional continuous-time model into an equivalent infinite-dimensional discrete-time model, e.g., an infinite-dimensional difference equation. Thus, the next step entails mapping of the equivalent infinite-dimensional discrete-time model to an essentially-equivalent infinite-dimensional discrete-time model; this is performed in Chapter IV. The work in Section 3.4 parallels that of Maybeck [129] for finite-dimensional systems, and hence the idea is not original, but the transformation process for infinite-dimensional systems is a contribution resulting from this research. That is, the linkage between the continuous-time dynamics model given in Definition 79 and the discrete-time dynamics model of Definition 80 is new even though it “looks” nearly identical to the finite-dimensional work as exemplified in Maybeck [129] — see Theorems 86 and 88 in particular.

The purpose of the extended example problem given in Chapter IV is to develop a method for using the ISKF in an MMAE to estimate the temperature profile along a slender cylindrical rod; the resulting structure realized by the MATLAB programming environment is the approximate infinite-dimensional MMAE (AIMMAE). By approximating the temperature (state) function, we are able to use the infinite-dimensional structural and statistical components of the ISKF without further approximations. Using the results of Section 3.4, we transformed the infinite-dimensional continuous-time system model into an equivalent infinite-dimensional

discrete-time model. Next, an essentially-equivalent finite-dimensional discrete-time model was derived by approximating the state temperature function using a finite number of terms from a Fourier series expansion and then determining the resulting form of the ISKF components. This approximation technique gives rise to a sampled-data Kalman filter used to generate an optimal estimate of a predetermined finite number of the coefficients associated with the Fourier series expansion of the true state temperature function.

In Chapter V we found that the AIMMAE was quite capable of estimating the state in a variety of uncertain noise environments as well as performing the task of system identification. It was not an original goal of this work to perform the system identification task, however, it accomplishes the task quite well.

6.3 Recommendations

There are several main threads of inquiry that warrant considerable more attention than given in this dissertation. Characterizing the ISKF, expanding the class of problems that the ISKF can be filter, improving the state function approximation, and the employment of moving-bank MMAE structures represent interesting areas for further research.

While the ISKF was fully developed, we have made no attempt to characterize the *controllability* (the property of being able to steer the system between two arbitrary points in the state space) or *observability* (the property of being able to determine the initial state uniquely using only the knowledge of the output) of the infinite-dimensional dynamics and measurement models upon which estimator was based or the *stability* of the ISKF itself. See texts by Curtain on infinite-dimensional system theory for a good discussion regarding these topics [38, 39].

Another area for exploration would be to expand the class of infinite-dimensional problems to include two-parameter semigroups of bounded linear operators; see, for example, Pazy [160].

In transforming the infinite-dimensional continuous-time system model into an equivalent infinite-dimensional discrete-time model, we experienced no loss of information. In contrast, the quality of essentially-equivalent finite-dimensional discrete-time model is wholly dependent on the method and subspace used to model the state function, since the structural and statistical components of the ISKF were developed in concert with the approximation chosen for the state function. Thus, the performance of the ISKF could be enhanced by optimizing the manner in which the state function is approximated.

In Chapter II we introduced several moving-bank structures. Depending on the application, one of these might improve state and/or parameter estimation performance relative to the fixed-bank MMAE used in this research.

Bibliography

1. *Britannica 2001 Standard Edition CD-ROM*. Encyclopædia Britannica, Inc., 2001.
2. Ackerson, Guy A. and K. S. Fu. "On State Estimation in Switching Environments," *IEEE Transactions on Automatic Control*, AC-15(1):10–17 (February 1970).
3. Anderson, Brian D. O. and John B. Moore. *Optimal Filtering*. Information and System Sciences, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1979.
4. Andersson, P. "Adaptive Forgetting in Recursive Identification through Multiple Models," *International Journal of Control*, 42:1175–1193 (November 1985).
5. Anton, Howard. *Calculus* (Second Edition). New York: John Wiley & Sons, Inc., 1984.
6. Antoulas, Anthanasios Constantinos, editor. *Mathematical System Theory: The Influence of Rudolf Emil Kalman*. Berlin: Springer-Verlag Berlin Heidelberg, 1991.
7. Apostol, Tom M. *Mathematical Analysis* (Second Edition). Reading, Massachusetts: Addison-Wesley Publishing Company, 1974.
8. Arfken, George. *Mathematical Methods for Physicists* (Third Edition). San Diego, California: Academic Press, Inc., 1985.
9. Åström, K. J. and P. Eykhoff. "System Identification—A Survey," *Automatica*, 7(2):123–162 (March 1971).
10. Athans, Michael, et al. "The Stochastic Control of the F-8C Aircraft Using a Multiple Model Adaptive Control (MMAC) Method—Part I: Equilibrium Flight," *IEEE Transactions on Automatic Control*, AC-22(5):768–780 (October 1977).
11. Athans, Michael and C. B. Chang. *Adaptive Estimation and Parameter Identification Using a Multiple Model Estimation Algorithm*. Technical Report ESD-TR-76-184, Lexington, Massachusetts: Lincoln Laboratory, Massachusetts Institute of Technology, June 1976. Technical Note 1976-28.
12. Başar, Tamer, editor. *Control Theory: Twenty-Five Seminal Papers (1932-1981)*. New York: IEEE Press, 2001.
13. Bar-Shalom, Yaakov and Thomas E. Fortmann. *Tracking and Data Association*. Mathematics in Science and Engineering, 179, Boston, Massachusetts: Academic Press, Inc., 1988.
14. Bar-Shalom, Yaakov, et al. *Estimation with Applications to Tracking and Navigation*. New York: John Wiley & Sons, Inc., 2001.

15. Baram, Yoram and Nils R. Sandell, Jr. "Consistent Estimation of Finite Parameter Sets with Application to Linear Systems Identification," *IEEE Transactions on Automatic Control*, AC-23(3):451–454 (June 1978).
16. Baram, Yoram and Nils R. Sandell, Jr. "An Information Theoretic Approach to Dynamic System Modelling and Identification," *IEEE Transactions on Automatic Control*, AC-23(1):61–66 (February 1978).
17. Barham, P. M. and D. E. Humphries. "Derivation of the Kalman Filtering Equations from Elementary Statistical Principles." *Theory and Applications of Kalman Filtering AGARDograph no. 139*, edited by Cornelius T. Leondes. 43–49. London: North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development and Technical Editing and Reproduction Ltd., February 1970.
18. Bensoussan, Alain, et al. *Representation and Control of Infinite Dimensional Systems, I*. Systems & Control: Foundations & Applications, Boston, Massachusetts: Birkhäuser, 1992.
19. Bensoussan, Alain, et al. *Representation and Control of Infinite Dimensional Systems, II*. Systems & Control: Foundations & Applications, Boston, Massachusetts: Birkhäuser, 1993.
20. Berg, Paul W. and James L. McGregor. *Elementary Partial Differential Equations*. Oakland, California: Holden-Day, Inc., 1966.
21. Billingsley, Patrick. *Probability and Measure* (Third Edition). New York: John Wiley & Sons, Inc., 1991.
22. Blom, H. A. P. "An Efficient Filter for Abruptly Changing Systems," *Proceedings of the 23th IEEE Conference on Decision and Control*, 656–658 (December 1984).
23. Blom, Henk A. P. and Yaakov Bar-Shalom. "The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients," *IEEE Transactions on Automatic Control*, 33(8):780–783 (August 1988).
24. Bobrowski, Adam. *Functional Analysis for Probability and Stochastic Processes*. Cambridge, United Kingdom: Cambridge University Press, 2005.
25. Borowski, Ephraim J. and Jonathan M. Borwein. *The Harper Collins Dictionary of Mathematics*. New York: HarperPerennial, 1991.
26. Boyce, William E. and Richard C. DiPrima. *Elementary Differential Equations and Boundary Value Problems*. New York: John Wiley & Sons, Inc., 1965.
27. Boyer, William H., editor. *CRC Standard Mathematical Tables* (28 Edition). Boca Raton, Florida: CRC Press, Inc., 1987.

28. Bracewell, Ronald N. *The Fourier Transform and Its Applications* (Third Edition). Boston, Massachusetts: The McGraw-Hill Companies, Inc., 2000.
29. Brammer, Karl and Gerard Siffing. *Kalman-Bucy Filters*. Norwood, Massachusetts: Artech House, Inc., 1989.
30. Brenner, Susanne C. and L. Ridgway Scott. *The Mathematical Theory of Finite Element Methods* (Second Edition). Texts in Applied Mathematics, 15, New York: Springer-Verlag New York, Inc., 2002.
31. Bryson, Jr., Arthur E. and Yu-Chi Ho. *Applied Optimal Control*. Washington, D.C.: Hemisphere Publishing Corp., Halsted Press, and John Wiley & Sons, Inc., 1975.
32. Bucy, Richard S. and Peter D. Joseph. *Filtering for Stochastic Processes with Applications to Guidance*. New York: John Wiley & Sons, Inc., 1968.
33. Catlin, Donald E. *Estimation, Control, and the Discrete Kalman Filter*. Applied Mathematical Sciences, 71, New York: Springer-Verlag New York, Inc., 1989.
34. Chang, C. B. and Michael Athans. "State Estimation for Discrete Systems with Switching Parameters," *IEEE Transactions on Aerospace and Electronic Systems*, AES-14(4):418–424 (May 1978).
35. Clark, Curtis Steven. *Multiple Model Adaptive Estimation and Control Redistribution Performance on the VISTA F-16 During Partial Actuator Impairments*. MS thesis, AFIT/GE/ENG/97D-23, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1997.
36. Conway, John B. *A Course in Functional Analysis*. Graduate Texts in Mathematics, 96, New York: Springer-Verlag New York, Inc., 1990.
37. Conway, John B. *The Theory of Subnormal Operators*. Mathematical Surveys and Monographs, 36, Providence, Rhode Island: American Mathematical Society, 1991.
38. Curtain, Ruth F. and Anthony J. Pritchard. *Infinite Dimensional Linear Systems Theory*. Lecture Notes in Control and Information Sciences, 8, Berlin: Springer-Verlag, 1978.
39. Curtain, Ruth F. and Hans J. Zwart. *An Introduction to Infinite-Dimensional Linear System Theory*. Texts in Applied Mathematics, 21, New York: Springer-Verlag New York, Inc., 1995.
40. Da Prato, Giuseppe and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and Its Applications, 45, Cambridge, United Kingdom: Cambridge University Press, 1998.

41. Dasgupta, S. and L. C. Westphal. "Convergence of Partitioned Adaptive Filters for Systems with Unknown Biases," *IEEE Transactions on Automatic Control*, AC-28(5):614–615 (May 1983).
42. Doob, J. L. *Stochastic Processes*. New York: John Wiley & Sons, Inc., 1953.
43. Dorny, C. Nelson. *A Vector Space Approach to Models and Optimization*. New York: John Wiley & Sons, Inc., 1975.
44. DuChateau, Paul and David Zachmann. *Applied Partial Differential Equations*. New York: Harper & Row, Publishers, Inc., 1989.
45. Durrett, Richard. *Probability: Theory and Examples* (Third Edition). Duxbury Advanced Series, Belmont, California: Thompson Learning Inc., Brooks/Cole Publishing Co., 2005.
46. Eide, Peter K. and Peter S. Maybeck. "An MMAE Failure Detection System for the F-16," *IEEE Transactions on Aerospace and Electronic Systems*, 32(3):1125–1136 (July 1996).
47. Eide, Peter Keith. *Implementation and Demonstration of a Multiple Model Adaptive Estimation Failure Detection System for the F-16*. MS thesis, AFIT/GE/ENG/94D-06, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1994.
48. Engel, Klaus-Jochen and Rainer Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Graduate Texts in Mathematics, 194, New York: Springer-Verlag New York, Inc., 2000.
49. Erickson, John W. *Multiple Model Adaptive Estimation (MMAE) Applied to the Ultra-Tightly Coupled IMU/GPS Methodology*. PhD dissertation, AFIT/DS/ENG/01-03, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, September 2001.
50. Erickson, John W., et al. "Multipath-Adaptive GPS/INS Receiver," *IEEE Transactions on Aerospace and Electronic Systems*, 41(4):645–657 (April 2005).
51. Falb, Peter L. "Infinite-Dimensional Filtering: The Kalman-Bucy Filter in Hilbert Space," *Information and Control*, 11(1/2):102–137 (July-August 1967).
52. Farlow, Stanley J. *Partial Differential Equations for Scientists and Engineers*. New York: John Wiley & Sons, Inc., 1982.
53. Filios, Paul G. *Moving-Bank Multiple Model Adaptive Algorithms Applied to Flexible Spacecraft Control*. MS thesis, AFIT/GE/ENG/85D-14, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1985.
54. Fisher, Kenneth A. *Multiple Model Adaptive Estimation Using Filter Spawning*. MS thesis, AFIT/EN/ENG/99M-09, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, March 1999.

55. Fisher, Kenneth A. and Peter S. Maybeck. "Multiple Model Adaptive Estimation with Filter Spawning," *Proceedings of the 2000 American Control Conference*, 4:2326–2331 (June 2000).
56. Fisher, Kenneth A. and Peter S. Maybeck. "Multiple Model Adaptive Estimation with Filter Spawning," *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):755–768 (July 2002).
57. Fitch, James A. and Peter S. Maybeck. "Multiple Model Adaptive Control of a Large Flexible Space Structure with Purposeful Dither for Enhanced Identifiability," *Proceedings of the 33rd IEEE Conference on Decision and Control*, 2904–2909 (December 1994).
58. Fitch, James Alan. *Multiple Model Adaptive Control of a Large Flexible Space Structure with Purposeful Dither for Enhanced Identifiability*. MS thesis, AFIT/EN/ENG/93D–01, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1993.
59. Fry, C. M. and A. P. Sage. "On Hierarchical Structure Adaptation and Systems Identification," *International Journal of Control*, 20(3):433–452 (September 1974).
60. Gelfand, I. M. and S. V. Fomin. *Calculus of Variations*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963.
61. Gershenfeld, Neil A. *The Nature of Mathematical Modeling*. Cambridge, United Kingdom: Cambridge University Press, 2002.
62. Gockenbach, Mark S. *Partial Differential Equations: Analytical and Numerical Methods*. Philadelphia: Society for Industrial and Applied Mathematics, 2002.
63. Goodson, R. Eugene and Michael P. Polis. "Identification of Parameters in Distributed Systems." *Distributed Parameter Systems: Identification, Estimation, and Control*, Control and Systems Theory, 6, edited by W. Harmon Ray and Demetrios G. Lainiotis. 47–133. New York: Marcel Dekker, Inc., 1978.
64. Griffen, Gordon C. and Peter S. Maybeck. "MMAE/MMAC Control for Bending with Multiple Uncertain Parameters," *Proceedings of the 34th IEEE Conference on Decision and Control*, 1153–1158 (December 1995).
65. Griffen, Gordon C. and Peter S. Maybeck. "MMAE/MMAC Control for Bending with Multiple Uncertain Parameters," *IEEE Transactions on Aerospace and Electronic Systems*, 33(3):903–912 (July 1997).
66. Grigoriu, Mircea. *Stochastic Calculus: Applications in Science and Engineering*. Boston, Massachusetts: Birkhäuser, 2002.
67. Gustafson, John A. and Peter S. Maybeck. "Control of a Large Flexible Space Structure with Moving-Bank Multiple Model Adaptive Algorithms," *Proceed-*

- ings of the 31st IEEE Conference on Decision and Control*, 1273–1278 (December 1992).
68. Gustafson, John A. and Peter S. Maybeck. “Flexible Spacestructure Control via Moving-Bank Multiple Model Algorithms,” *IEEE Transactions on Aerospace and Electronic Systems*, 30(3):750–757 (July 1994).
 69. Gustafson, John Arthur. *Control of a Large Flexible Space Structure Using Multiple Model Adaptive Algorithms*. MS thesis, AFIT/GE/ENG/91D–22, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1991.
 70. Hanlon, Peter D. *Failure Identification Using Multiple Model Adaptive Estimation for the Lambda Flight Vehicle*. MS thesis, AFIT/GE/ENG/92D–19, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1992.
 71. Hanlon, Peter D. *Practical Implementation of Multiple Model Adaptive Estimation Using Neyman-Pearson Based Hypothesis Testing and Spectral Estimation Tools*. PhD dissertation, AFIT/DS/ENG/96-07, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1996.
 72. Hanlon, Peter D. and Peter S. Maybeck. “Interrelationship of Single-Filter and Multiple-Model Adaptive Algorithms,” *IEEE Transactions on Aerospace and Electronic Systems*, 34(3):934–946 (July 1998).
 73. Hanlon, Peter D. and Peter S. Maybeck. “Multiple-Model Adaptive Estimation Using Residual Correlation Kalman Filter Bank,” *IEEE Transactions on Aerospace and Electronic Systems*, 36(2):393–406 (April 2000).
 74. Harambasic, Jr., Louis Jasper. *Tracking and Control of a Neutral Particle Beam Using Multiple Model Adaptive Meer Filter*. MS thesis, AFIT/GE/ENG/87D–23, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1987.
 75. Hawkes, Richard M. and John B. Moore. “Performance Bounds for Adaptive Estimation,” *Proceedings of the IEEE*, 64(8):1143–1150 (August 1976).
 76. Hawkes, Richard M. and John B. Moore. “Performance of Bayesian Parameter Estimation for Linear Signal Models,” *IEEE Transactions on Automatic Control*, AC-21(4):523–527 (August 1976).
 77. Haykin, Simon. *Adaptive Filter Theory* (Third Edition). Upper Saddle River, New Jersey: Prentice-Hall, Inc., 1996.
 78. Heffes, H. “The Effect of Erroneous Models on the Kalman Filter Response,” *IEEE Transactions on Automatic Control*, AC-11:541–543 (July 1966).

79. Henderson, Paul E. *Development and Testing of a Multiple Filter Approach for Precise DGPS Positioning and Carrier-Phase Ambiguity Resolution*. MS thesis, AFIT/GE/ENG/01M-15, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, March 2001.
80. Henderson, Paul E. and John F. Raquet. "Development and Testing of a Multiple Filter Approach for Precise DGPS Positioning and Carrier-Phase Ambiguity Resolution," *Proceedings of the ION National Technical Meeting*, 806-815 (January 2000).
81. Hentz, Karl P. *Feasibility Analysis of Moving Bank Multiple Model Adaptive Estimation and Control Algorithms*. MS thesis, AFIT/GE/ENG/84D-32, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1984.
82. Hillborn, Jr., Charles G. and Demetrios G. Lainiotis. "Optimal Estimation in the Presence of Unknown Parameters," *IEEE Transactions on Systems Science and Cybernetics*, SSC-5(1):38-43 (January 1969).
83. Hille, Einar and Ralph S. Phillips. *Functional Analysis and Semi-Groups*. American Mathematical Society Colloquium Publications, 31, Providence, Rhode Island: American Mathematical Society, 1957.
84. Hoffman, Gregory S. *A Novel Electrocardiogram Segmentation Algorithm Using a Multiple Model Adaptive Estimator*. MS thesis, AFIT/GE/ENG/02M-10, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, March 2002.
85. Hoffman, Kenneth and Ray Kunze. *Linear Algebra* (Second Edition). Upper Saddle River, New Jersey: Prentice-Hall, Inc., 1971.
86. Hogg, Robert V. and Allen T. Craig. *Introduction to Mathematical Statistics* (Fifth Edition). Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1995.
87. Holden, Helge, et al. *Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach*. Boston, Massachusetts: Birkhäuser, 1996.
88. Howard, Randall Barry. *Confidence Interval Estimation for Output of Discrete-Event Simulations Using the Kalman Filter*. MS thesis, AFIT/GOR/ENS/92M-15, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, March 1992.
89. Hughes, Thomas J. R. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Mineola, New York: Dover Publications, Inc., 2000. Originally published: Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1987.
90. Jacobs, David and Alec Porter. *MMAE Final Project*. EENG 768 Stochastic Estimation and Control, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, September 2002.

91. Jazwinski, Andrew H. *Stochastic Processes and Filtering Theory*. Mathematics in Science and Engineering, 64, New York: Academic Press, Inc., 1970.
92. Jilkov, Vesselin P. and X. Rong Li. "Online Bayesian Estimation of Transition Probabilities for Markovian Jump Systems," *IEEE Transactions on Signal Processing*, 52(6):1620–1630 (June 2004).
93. Johnson, Bruce A. *Stochastic Adaptive Particle Beam Tracker Using Meer Filter Feedback*. MS thesis, AFIT/GE/ENG/86D–27, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1986.
94. Kailath, Thomas. "An Innovations Approach to Least-Squares Estimation Part I: Linear Filtering in Additive White Noise," *IEEE Transactions on Automatic Control*, AC-13(6):646–655 (December 1968).
95. Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME, Journal of Basic Engineering*, 82D:35–45 (March 1960).
96. Kalman, R. E. and R. S. Bucy. "New Results in Linear Filtering and Prediction Theory," *Transactions of the ASME, Journal of Basic Engineering*, 83D(1):95–108 (March 1961).
97. Kalman, Rudolf E., et al. *Topics in Mathematical System Theory*. International Series in Pure and Applied Mathematics, New York: McGraw-Hill, Inc., 1969.
98. Karnick, Drew A. *Moving-Bank Multiple Model Adaptive Estimation Applied to Flexible Spacestructure Control*. MS thesis, AFIT/GE/ENG/86D–41, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1986.
99. Kay, Steven M. *Modern Spectral Estimation*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1988.
100. Kay, Steven M. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, New Jersey: Prentice-Hall, Inc., 1993.
101. Kay, Steven M. *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, New Jersey: Prentice-Hall, Inc., 1998.
102. Koivo, H. N. and A. J. Koivo. "Control and Estimation of Systems with Time Delays." *Distributed Parameter Systems: Identification, Estimation, and Control*, Control and Systems Theory, 6, edited by W. Harmon Ray and Demetrios G. Lainiotis. 249–320. New York: Marcel Dekker, Inc., 1978.
103. Kolmanovskii, V. and A. Myshkis. *Introduction to the Theory and Applications of Functional Differential Equations*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1999.

104. Krishnan, Venkatarama. *Nonlinear Filtering and Smoothing*. New York: John Wiley & Sons, Inc., 1984. Republished: Mineola, New York: Dover Publications, Inc, 2005.
105. Kyger, David W. *Reducing Lag in Virtual Displays Using Multiple Model Adaptive Estimation*. MS thesis, AFIT/GE/ENG/95D-11, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1995.
106. Kyger, David W. and Peter S. Maybeck. "Reducing Lag in Virtual Displays Using Multiple-Model Adaptive Estimation," *Proceedings of the 1997 American Control Conference*, 4:2536-2541 (June 1997).
107. Lainiotis, Demetrios G. "Optimal Adaptive Estimation: Structure and Parameter Adaptation," *IEEE Transactions on Automatic Control*, AC-16(2):160-170 (April 1971).
108. Lainiotis, Demetrios G. "Partitioning: A Unifying Framework for Adaptive Systems, I: Estimation," *Proceedings of the IEEE*, 64(8):1126-1143 (August 1976).
109. Lainiotis, Demetrios G. "Partitioning: A Unifying Framework for Adaptive Systems, II: Control," *Proceedings of the IEEE*, 64(8):1182-1198 (August 1976).
110. Lamb, P. R. and L. C. Westphal. "Simplex-Directed Partitioned Adaptive Filters," *International Journal of Control*, 30(4):617-627 (October 1979).
111. Landau, Rubin H. and Manuel José Páez Mejía. *Computational Physics: Problem Solving with Computers*. New York: John Wiley & Sons, Inc., 1997.
112. Larson, Craig and Jason Williams. *Monte Carlo Model Selection — MMAE Final Project*. EENG 768 Stochastic Estimation and Control, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, September 2002.
113. Lashlee, Robert W. *Moving-Bank Multiple Model Adaptive Estimation Applied to Flexible Spacestructure Control*. MS thesis, AFIT/GE/ENG/87D-36, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1987.
114. Lashlee, Jr., Robert W. and Peter S. Maybeck. "Spacestructure Control Using Moving Bank Multiple Model Adaptive Estimation," *Proceedings of the 27th IEEE Conference on Decision and Control*, 712-717 (December 1988).
115. Lax, Peter D. *Functional Analysis*. New York: John Wiley & Sons, Inc., 2002.
116. Leland, Robert. "Approximate Maximum Likelihood Parameter Estimates for Stochastic Distributed Parameter Systems," *Proceedings of the 1997 American Control Conference*, 4:3693-3697 (June 1997).

117. Li, X. Rong. "Multiple-Model Estimation with Variable Structure — Part II: Model-Set Adaptation," *IEEE Transactions on Automatic Control*, 45(11):2047–2060 (November 2000).
118. Li, X. Rong, et al. "Multiple-Model Estimation with Variable Structure: Model-Group Switching Algorithm," *Proceedings of the 36th IEEE Conference on Decision and Control*, 3114–3119 (December 1997).
119. Li, X. Rong, et al. "General Model-Set Design Methods for Multiple-Model Approach," *IEEE Transactions on Automatic Control*, 50(9):1260–1276 (September 2005).
120. Li, Xiao-Rong and Yaakov Bar-Shalom. "Multiple-Model Estimation with Variable Structure," *IEEE Transactions on Automatic Control*, 41(4):478–493 (April 1996).
121. Ljung, Lennart. *System Identification: Theory for the User* (Second Edition). Upper Saddle River, New Jersey: Prentice Hall PTR, 1998.
122. Luenberger, David G. *Optimization by Vector Space Methods*. New York: John Wiley & Sons, Inc., 1969.
123. Lund, Eivin J. *On-line Discrimination and Estimation in Multiple Regime Systems*. PhD dissertation, Division of Engineering Cybernetics, The Norwegian Institute of Technology, University of Trondheim, Trondheim, Norway, June 1992.
124. Lund, Eivin J., et al. "Multiple Model Estimation with Inter-Residual Distance Feedback," *Modeling, Identification, and Control*, 13(3):127–140 (1992).
125. Magill, D. T. "Optimal Adaptive Estimation of Sampled Stochastic Processes," *IEEE Transactions on Automatic Control*, AC-10(4):434–439 (October 1965).
126. Marsden, Jerrold E. and Anthony J. Tromba. *Vector Calculus*. New York: W. H. Freeman and Company, 1988.
127. Martin, James F., et al. "Multiple-Model Adaptive Control of Blood Pressure using Sodium Nitroprusside," *IEEE Transactions on Biomedical Engineering*, BME-34(8):603–611 (August 1987).
128. Matthes, Jr., James R. *Model Selection for the Multiple Model Adaptive Algorithm for In-Flight Simulation*. MS thesis, AFIT/GE/ENG/87D–40, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1987.
129. Maybeck, Peter S. *Stochastic Models, Estimation, and Control*, 1. Mathematics in Science and Engineering, 141. New York: Academic Press, Inc., 1979. Republished: Arlington, Virginia: Navtech Press, 1994.

130. Maybeck, Peter S. *Stochastic Models, Estimation, and Control*, 2. Mathematics in Science and Engineering, 141. New York: Academic Press, Inc., 1982. Republished: Arlington, Virginia: Navtech Press, 1994.
131. Maybeck, Peter S. *Stochastic Models, Estimation, and Control*, 3. Mathematics in Science and Engineering, 141. New York: Academic Press, Inc., 1982. Republished: Arlington, Virginia: Navtech Press, 2001.
132. Maybeck, Peter S. "Moving-Bank Multiple Model Adaptive Estimation and Control Algorithms: An Evaluation." *Control and Dynamic Systems: Advances in Theory and Applications* 31, edited by Cornelius T. Leondes. 1–31. New York: Academic Press, Inc., 1989.
133. Maybeck, Peter S. *Multiple Model Adaptive Estimation*. EENG 768 Stochastic Estimation and Control, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, July 2001.
134. Maybeck, Peter S. *2002 MMAE Matlab Project Software*. EENG 768 Stochastic Estimation and Control, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, July 2002.
135. Maybeck, Peter S. and Peter D. Hanlon. "Performance Enhancement of a Multiple Model Adaptive Estimator," *IEEE Transactions on Aerospace and Electronic Systems*, 31(4):1240–1254 (October 1995).
136. Maybeck, Peter S. and Karl P. Hentz. "Investigation of Moving-Bank Multiple Model Adaptive Algorithms," *AIAA Journal of Guidance, Control, and Dynamics*, 10(1):90–96 (January-February 1987).
137. Maybeck, Peter S. and Brian D. Smith. "Multiple Model Tracker Based on Gaussian Mixture Reduction for Maneuvering Targets in Clutter," *Proceedings of the 8th International Society of Information Fusion Conference* (July 2005).
138. Maybeck, Peter S. and Richard D. Stevens. "Reconfigurable Flight Control via Multiple Model Adaptive Control Methods," *Proceedings of the 29th IEEE Conference on Decision and Control*, 3351–3356 (December 1990).
139. Maybeck, Peter S. and Richard D. Stevens. "Reconfigurable Flight Control via Multiple Model Adaptive Control Methods," *IEEE Transactions on Aerospace and Electronic Systems*, 27(3):470–480 (May 1991).
140. Mazor, Efim, et al. "Interacting Multiple Model Methods in Target Tracking: A Survey," *IEEE Transactions on Aerospace and Electronic Systems*, 34(1):103–123 (January 1998).
141. McGarty, Terrence P. *Stochastic Systems and State Estimation*. New York: John Wiley & Sons, Inc., 1974.

142. Meer, David E. *Multiple Model Adaptive Estimation for Space-Time Point Process Observations*. PhD dissertation, AFIT/DS/EE/82-2, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, September 1982.
143. Mehra, Raman K. "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE Transactions on Automatic Control*, AC-15(2):175–184 (April 1970).
144. Meidel, K. W., et al. "GPS Signal Jamming Mitigation through Multiple Model Adaptive Estimation Applied to Ultra-Tightly Coupled GPS/INS Architecture," *Proceedings of the Institute of Navigation (ION) Global Navigation Satellite System (GNSS) Conference* (September 2005).
145. Menke, Timothy E. *Multiple Model Adaptive Estimation Applied to the VISTA F-16 with Actuator and Sensor Failures*. MS thesis, AFIT/GE/ENG/92J-01, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, June 1992.
146. Menke, Timothy E. and Peter S. Maybeck. "Multiple Model Adaptive Estimation Applied to the VISTA F-16 Flight Control System with Actuator and Sensor Failures," *Proceedings of the IEEE National Aerospace and Electronics Conference*, 441–448 (May 1992).
147. Menke, Timothy E. and Peter S. Maybeck. "Sensor/Actuator failure Detection in the VISTA F-16 by Multiple Model Adaptive Estimation," *IEEE Transactions on Aerospace and Electronic Systems*, 31(4):1218–1229 (October 1995).
148. Miller, Mikel M. and Peter S. Maybeck. "Modified Multiple Model Adaptive Estimation (M^3AE) for Accurate Navigation and Detection of Electromagnetic Interference to the GPS Signal," *Proceedings of the ION GPS-98 Conference*, 895–904 (September 1998).
149. Miller, Mikel Mark. *Modified Multiple Model Adaptive Estimation (M^3AE) for Simultaneous Parameter and State Estimation*. PhD dissertation, AFIT/DS/ENG/98-02, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, March 1998.
150. Moose, Richard L. "An Adaptive State Estimation Solution to the Maneuvering Target Problem," *IEEE Transactions on Automatic Control*, AC-20(3):359–362 (June 1975).
151. Moose, Richard L. and P. P. Wang. "An Adaptive Estimator with Learning for a Plant Containing Semi-Markov Switching Parameters," *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3:277–281 (May 1973).
152. Mosle, III, William Boardman. *Failure Detection, Isolation, and Recovery in an Integrated Navigation System*. MS thesis, AFIT/GE/ENG/93D-28, School of

- Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1993.
153. Muravez, Randall Joseph. *Multiple Model Adaptive Estimation and Prediction with the Harmonically Balanced Kalman Filter Bank*. MS thesis, California State Polytechnic University, Pomona, California, 1989.
 154. Naylor, Arch W. and George R. Sell. *Linear Operator Theory in Engineering and Science*. Applied Mathematical Sciences, 40, New York: Springer-Verlag, 1982.
 155. Ogata, Katsuhiko. *Modern Control Engineering* (Third Edition). Upper Saddle River, New Jersey: Prentice-Hall, Inc., 1997.
 156. Ormsby, Charles D. *Generalized Residual Multiple Model Adaptive Estimation of Parameters and States*. PhD dissertation, AFIT/DS/ENG/03-08, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, October 2003.
 157. Ormsby, Charles D., et al. "A Generalized Multiple Model Adaptive Estimator for Parameter and State Estimation," *Proceedings of the 59th Annual Meeting of the Institute of Navigation (ION)* (June 2003).
 158. Oxley, Mark E. *Linear Functional Analysis*. MATH 705 Class Notes, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, September 2001.
 159. Papoulis, Athanasios. *Probability, Random Variables, and Stochastic Processes* (Third Edition). New York: McGraw-Hill, Inc., 1991.
 160. Pazy, Amnon. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Applied Mathematical Sciences, 44, New York: Springer-Verlag New York, Inc., 1983.
 161. Phillipson, Gordon A. *Identification of Distributed Systems*. Modern Analytic and Computational Methods in Science and Mathematics, 29, New York: American Elsevier Publishing Company, Inc., 1971.
 162. Polya, George. *How to Solve It* (Second Edition). Princeton, New Jersey: Princeton University Press, 1973.
 163. Ray, W. Harmon and Demetrios G. Lainiotis, editors. *Distributed Parameter Systems: Identification, Estimation, and Control*. Control and Systems Theory, 6, New York: Marcel Dekker, Inc., 1978.
 164. Reid, J. Gary. *Linear System Fundamentals: Continuous and Discrete, Classic and Modern*. New York: McGraw-Hill, Inc., 1983.
 165. Retherford, J. R. *Hilbert Space: Compact Operators and the Trace Theorem*. London Mathematical Society Student Texts, 27, Cambridge, United Kingdom: Cambridge University Press, 1993.

166. Royden, H. L. *Real Analysis* (Third Edition). New York: Macmillan Publishing Company, 1988.
167. Sage, Andrew and James Melsa. *System Identification*. Mathematics in Science and Engineering, 80, New York: Academic Press, Inc., 1971.
168. Sanders, Donald H. *Statistics: A First Course* (Fifth Edition). Boston, Massachusetts: McGraw-Hill, Inc., 1995.
169. Sangsuk-Iam, Suwanchai and Thomas E. Bullock. "Analysis of Discrete-Time Kalman Filtering Under Incorrect Noise Covariances," *IEEE Transactions on Automatic Control*, 35(12):1304–1309 (December 1990).
170. Scharf, Louis L. *Statistical Signal Processing*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1991.
171. Schiller, Gregory J. and Peter S. Maybeck. "Control of a Large Space Structure Using MMAE/MMAC Techniques," *IEEE Transactions on Aerospace and Electronic Systems*, 33(4):1122–1131 (October 1997).
172. Schiller, Gregory John. *Control of a Large Space Structure Using Multiple Model Adaptive Estimation and Control Techniques*. MS thesis, AFIT/EN/ENG/93D-02, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1993.
173. Schott, Kevin D. and B. Wayne Bequette. "Multiple Model Adaptive Control." *Multiple Model Approaches to Modelling and Control* edited by Roderick Murray-Smith and Tor Arne Johansen, Bristol, Pennsylvania: Taylor & Francis, Inc., 1997.
174. Sengbush, R. L. and Demetrios G. Lainiotis. "Simplified Parameter Quantization Procedure for Adaptive Estimation," *IEEE Transactions on Automatic Control*, AC-14(4):424–425 (August 1969).
175. Sheldon, Stuart N. *An Optimal Parameter Discretization Strategy for Multiple Model Adaptive Estimation and Control*. PhD dissertation, AFIT/DS/ENG/89-2, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1989.
176. Sheldon, Stuart N. and Peter S. Maybeck. "An Optimizing Design Strategy for Multiple Model Adaptive Estimation and Control," *Proceedings of the 29th IEEE Conference on Decision and Control*, 3522–3527 (December 1990).
177. Sheldon, Stuart N. and Peter S. Maybeck. "An Optimizing Design Strategy for Multiple Model Adaptive Estimation," *IEEE Transactions on Automatic Control*, 38(4):651–654 (April 1993).
178. Smith, Brian D. *Multiple Model Adaptive Estimator Target Tracker for Maneuvering Targets in Clutter*. MS thesis, AFIT/GE/EE/05-18, School of Engi-

- neering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, March 2005.
179. Smith, N. J. and A. P. Sage. "Hierarchical Structuring for System Identification," *Information Sciences*, 7:49–72 (1974).
 180. Snyder, Donald L. and Michael I. Miller. *Random POint Processes in Time and Space* (Second Edition). New York: Springer-Verlag New York, Inc., 1991.
 181. Sorenson, H. W. and A. R. Stubberud. "Linear Estimation Theory." *Theory and Applications of Kalman Filtering* AGARDograph no. 139, edited by Cornelius T. Leondes. 3–42. London: North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development and Technical Editing and Reproduction Ltd., February 1970.
 182. Sorenson, Harold W. *Parameter Estimation: Principles and Problems*. Control and Systems Theory, 9, New York: Marcel Dekker, Inc., 1980.
 183. Sorenson, Harold W., editor. *Kalman Filtering: Theory and Application*. New York: IEEE Press, 1985.
 184. Stavroulakis, Peter, editor. *Distributed Parameter Systems Theory, Part I: Control*. Benchmark Papers in Electrical Engineering and Computer Science, 26, Stroudsburg, Pennsylvania: Hutchinson Ross Publishing Company, 1983.
 185. Stavroulakis, Peter, editor. *Distributed Parameter Systems Theory, Part II: Estimation*. Benchmark Papers in Electrical Engineering and Computer Science, 26, Stroudsburg, Pennsylvania: Hutchinson Ross Publishing Company, 1983.
 186. Stepaniak, Michael J. and Peter S. Maybeck. "MMAE-Based Control Redistribution Applied to the VISTA F-16," *IEEE Transactions on Aerospace and Electronic Systems*, 34(4):1249–1260 (October 1998).
 187. Stepaniak, Michael Joseph. *Multiple Model Adaptive Control of the VISTA F-16*. MS thesis, AFIT/GE/ENG/95D–26, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1995.
 188. Stevens, Richard D. *Characterization of a Reconfigurable Multiple Model Adaptive Controller Using a STOL F-15 Model*. MS thesis, AFIT/GE/ENG/89D–52, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1989.
 189. Strang, Gilbert. *Linear Algebra and Its Applications* (Third Edition). Fort Worth, Texas: Harcourt Brace Jovanovich, Inc., 1988.
 190. Strang, Gilbert and Kai Borre. *Linear Algebra, Geodesy, and GPS*. Wellesley, Massachusetts: Wellesley-Cambridge Press, 1997.

191. Suizu, Robert I. *Enhanced Tracking of Airborne Targets Using Multiple Model Filtering Techniques for Adaptive Field of View Expansion*. MS thesis, AFIT/GE/ENG/84M-4, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, March 1984.
192. Sullivan, Joe H. and William H. Woodall. "Estimating Markov Transition Matrices Using Uncertain Observed States," *Stochastic Analysis and Applications*, 17(2):253–274 (1999).
193. Sworder, David D. and John E. Boyd. *Estimation Problems in Hybrid Systems*. Cambridge, United Kingdom: Cambridge University Press, 1999.
194. Taylor, Angus E. *General Theory of Functions and Integration*. New York: Blaisdell Publishing Company, 1965.
195. Tugnait, Jitendra K. "Comments on "State Estimation for Discrete Systems with Switching Parameters"," *IEEE Transactions on Aerospace and Electronic Systems*, AES-15(3):464 (May 1979).
196. Tugnait, Jitendra K. "Convergence Analysis of Partitioned Adaptive Estimators Under Continuous Parameter Uncertainty," *IEEE Transactions on Automatic Control*, AC-25(3):569–573 (June 1980).
197. Tugnait, Jitendra K. "Detection and Estimation for Abruptly Changing Systems," *Automatica*, 18(5):607–615 (September 1982).
198. Vasquez, Juan R. *New Algorithms for Moving-Bank Multiple Model Adaptive Estimation*. PhD dissertation, AFIT/DS/ENG/98-10, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, May 1998.
199. Vasquez, Juan R. and Peter S. Maybeck. "Density Algorithm Based Moving-Bank MMAE," *Proceedings of the 38th IEEE Conference on Decision and Control*, 4117–4122 (December 1999).
200. Vasquez, Juan R. and Peter S. Maybeck. "Enhanced Motion and Sizing of Bank in Moving-Bank MMAE," *Proceedings of the 1999 American Control Conference*, 1555–1562 (June 1999).
201. Vasquez, Juan R. and Peter S. Maybeck. "Enhanced Motion and Sizing of Bank in Moving-Bank MMAE," *IEEE Transactions on Aerospace and Electronic Systems*, 40(3):770–779 (July 2004).
202. Wald, Abraham. *Sequential Analysis*. New York: John Wiley & Sons, Inc., 1947.
203. Watanabe, Keigo and Spyros G. Tzafestas. "A Hierarchical Multiple Model Adaptive Control of Discrete-Time Stochastic Systems for Sensor and Actuator Uncertainties," *Automatica*, 26(5):875–886 (1990).

204. White, Nathan A., et al. "Detection of Interference/Jamming and Spoofing in a DGPS-Aided Inertial System," *IEEE Transactions on Aerospace and Electronic Systems*, 34(4):1208–1217 (October 1998).
205. White, Nathan A., et al. "MMAE Detection of Interference/Jamming and Spoofing in a DGPS-Aided Inertial System," *Proceedings of the ION GPS-98 Conference*, 905–914 (September 1998).
206. White, Nathan Alan. *MMAE Detection of Interference/Jamming and Spoofing in a DGPS-Aided Inertial System*. MS thesis, AFIT/GE/ENG/96D–21, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1996.
207. Whittle, P. *Prediction and Regulation by Linear Least-Square Methods*. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1963.
208. Wiener, Norbert. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: The Technology Press of the Massachusetts Institute of Technology and John Wiley & Sons, Inc., 1949.
209. Wiener, Norbert, et al. *Differential Space, Quantum Systems, and Prediction*. Cambridge, Massachusetts: M.I.T. Press, 1966.
210. Williams, Jason L. *Gaussian Mixture Reduction for Tracking Multiple Maneuvering targets in Clutter*. MS thesis, AFIT/GE/ENG/03–19, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, March 2003.
211. Willsky, Alan S. "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, 12:601–611 (December 1976).
212. Willsky, Alan S., et al. "Dynamic Model-Based Techniques for the Detection of Incidents on Freeways," *IEEE Transactions on Automatic Control*, AC-25(3):347–359 (June 1980).
213. Willsky, Alan S. and H. L. Jones. "A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems," *IEEE Transactions on Automatic Control*, AC-21(1):108–112 (February 1976).
214. Willsky, Alan S. and Harold L. Jones. "A Generalized Likelihood Ratio Approach to State Estimation in Linear Systems to Abrupt Changes," *Proceedings of the 1974 IEEE Conference on Decision and Control Including the 13th Symposium on Adaptive Processes*, 846–853 (November 1974).
215. Woodruff, Eileen A., et al. "A Model for the Design and Evaluation of Algorithm for Closed-Loop Cardiovascular Therapy," *IEEE Transactions on Biomedical Engineering*, 44(8):694–705 (August 1997).
216. Yosida, Kôzaku. *Functional Analysis* (Sixth Edition). A Series of Comprehensive Studies in Mathematics, 123, Berlin: Springer-Verlag, 1980.

217. Zadeh, Lofti A. and Charles A. Desoer. *Linear System Theory: The State Space Approach*. New York: McGraw-Hill Book Company, Inc., 1963.
218. Zicker, William L. *Pointing and Tracking of Particle Beams*. MS thesis, AFIT/GE/ENG/83D-73, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, December 1983.

Vita

Scott A. Sallberg was born and raised in Minnesota — the Land of 10,000 Lakes. He was graduated from Spring Lake Park High School in May 1984 and entered the United States Air Force (USAF) in September. Following graduation from the Basic Military Training School (BMTS) at Lackland Air Force Base (AFB), Texas and attendance at the Computer Operations Course held at Keesler AFB, Mississippi, Airman Sallberg was assigned to the TRIAD Computer Systems (TRICOMS), Headquarters Strategic Air Command (SAC), Offutt AFB, Nebraska, from January 1985 to December 1987. While at Offutt AFB, Airman Sallberg began his bachelors of science degree at the University of Nebraska-Omaha (UNO). In January 1988 he transferred to the University of Missouri-Rolla (UMR) under the sponsorship of the Airman's Education and Commissioning Program (AECPP). He was graduated, cum laude, in December 1990, having completed his Bachelors of Science in Electrical Engineering (BSEE) with minors in physics and mathematics. He was commissioned a Second Lieutenant at the Air Force Officer Training School (OTS), Medina Annex, Lackland AFB, Texas in June 1991. During his first assignment as an officer, Second Lieutenant Sallberg was an Aerodynamic Weapons Design Engineer in the Engineering Division at the National Air Intelligence Center [NAIC, formerly Foreign Technology Division (FTD)], Wright Patterson AFB, Ohio. After three years at NAIC, First Lieutenant Sallberg was selected to continue his formal education at the Air Force Institute of Technology (AFIT) in May 1994. He specialized in signal and optical information processing while pursuing the Masters of Science in Electrical Engineering (MSEE); he was graduated in December 1995. Capt Sallberg's follow-on assignment was with the Avionics Directorate of Wright Laboratory (WL), Wright Patterson AFB, Ohio; there he investigated and managed multi-spectral infrared detection and estimation application programs as a member of the Electro-Optics Division. Captain Sallberg served as an executive officer to the director of Avionics Directorate and then with Sensors Directorate of the Air Force

Research Laboratory (AFRL) from May 1997 until March 1998. Next, enroute to the Air Force Information Warfare Center (AFIWC) at the former Kelly AFB (now part of Lackland AFB), Texas, he attended Squadron Officer School (SOS) at Maxwell AFB, Alabama. At AFIWC, he was the Combat Support Database (CSDB) Production Manager; the database was used by the flight mission planning community to optimize route and ordinance delivery planning. Additionally, Mr. Sallberg was a statistics instructor for Park College at Kelly AFB in 1999. In August 2000, Captain Sallberg returned to AFIT to begin the Doctorate of Philosophy (PhD) in Electrical Engineering (specializing in stochastic estimation — a sub-specialty of the guidance and control discipline). After successfully defending his research, but prior to completing the dissertation document, Major Sallberg was transferred to Headquarters Air Force Operational Test and Evaluation Center (AFOTEC), Kirtland AFB, New Mexico in July 2004. Scott is retiring from active duty in July 2007 after nearly 23 years of service to our country and the USAF.

REPORT DOCUMENTATION PAGE				<i>Form Approved OMB No. 0704-0188</i>	
<small>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</small>					
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.					
1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)